UNIVERSITÄT SIEGEN

2019 WI Siegen

Tagungsband

23.02.2019 - 27.02.2019

Human Practice.

Digital Ecologies.

Our Future.

Thomas Ludwig und Volkmar Pipek (Hrsg.)

Human Practice.
Digital Ecologies.
Our Future.

14. Internationale Tagung Wirtschaftsinformatik (WI 2019)

Tagungsband

Vorwort der Tagungsleitung

Die Internationale Tagungsserie Wirtschaftsinformatik (WI 2019) ist mit ihrer 14. Ausgabe seit über 25 Jahren die größte Wirtschaftsinformatik-Konferenz im deutschsprachigen Raum und ihr Themenfeld hat nichts von seiner ökonomischen und gesellschaftlichen Brisanz verloren. Es scheint im Gegenteil so zu sein, dass erst langsam deutlich wird, wie konsequent Informationstechnologien jeden Aspekt unseres Alltags, ja unserer Existenz, umgestalten. Heute nennen wir diesen Umgestaltungsprozess „Digitalisierung“, und er entspricht bildlich gesehen einer Flamme, von der immer noch nicht klar sein mag, ob sie wärmt oder verbrennt, aber jedem wird klar, dass man sich damit auseinandersetzen muss.

Was bedeutet dies für die Wirtschaftsinformatik als Motor und wissenschaftlicher Unterbau dieses Umgestaltungsprozesses? Der Natur einer Wissenschaft entspricht es, sich nicht auf Ihren Erfolgen auszuruhen, sondern das Erreichte in seinen ökonomischen und gesellschaftlichen Effekten kritisch zu reflektieren, und nach neuen Herausforderungen zu suchen. Aus diesem Grund stellten wir die Konferenz 2019 in Siegen unter das Motto

„Human Practice. Digital Ecologies. Our Future.“

Damit bringen wir das komplexe Geflecht aus menschlicher Praxis und Praktik sowie ganzer digitalen Ökologie zu Ausdruck, das einen starken Einfluss auf unsere gemeinsame Zukunft und die Art wie wir arbeiten, aber vor allem auch leben werden, haben wird.

Mitten in diesem komplexen Geflecht der Digitalisierung erfahren wir, wie unsere Wissenschaftsdisziplin ihre Monopolstellung zu ökonomisch relevanten Digitalisierungsfragen in dem Maße verliert, in dem sich andere Disziplinen mit diesen Konzepten und Effekten stärker auseinandersetzen. Hier finden sich Beispiele von der Politologie über Jura, bis hin zur Verfahrenstechnik und zum Maschinenbau. „Unsere“ Zukunft ist eben nicht nur die Zukunft der Wirtschaftsinformatik an sich, sondern auch die Zukunft der Gestaltung ihrer Beziehung zu anderen Domänen und Wissenschaftsdisziplinen.

Dieses Zusammenwirken der verschiedenen Anwendungsbereiche haben wir versucht in 14 wissenschaftlichen Tracks widerzuspiegeln. Diese reichten von bereits etablierten Tracks wie Unternehmensmodellierung oder IT-Management bis zu den eher neueren Tracks wie Cyber-Physische Systeme, eHealth oder Umweltinformatik. Um die Wirtschaftsinformatik aus den verschiedenen gesellschaftlichen Bereichen zu betrachten, fokussierten die Keynotes der drei Tage auf die Bereiche Industrie (Herr Dr. Robert Zores von der REWE Digital GmbH und Frau Claudia Lensker von der Finanz Informatik GmbH & Co. KG), Wissenschaft (Prof. Dr. Gerhard Fischer von der

University of Colorado Boulder) und Verbraucher (Frau Helga Zander-Hayat von der Verbraucherzentrale NRW e.V.). Die zusätzlichen vier Panels zu Themen wie KI-basierte Digitale Assistenten, Arbeit 4.0 oder DSGVO untermauerten den Gedanken einer vernetzten Wirtschaftsinformatik.

Ergänzt wurde das wissenschaftliche Programm durch drei Special Tracks. Neben dem Praktiker- sowie Student-Track, gab es erstmalig auch einen expliziten Poster- und Demotrack, um vor allem auch NachwuchswissenschaftlerInnen die Möglichkeit zu geben, ihre Arbeiten zu präsentieren. Zusätzlich fand im Vorfeld der Konferenz das Doctorial Consortium, sowie der Workshoptag bestehend aus acht Workshops statt, in denen kontrovers über Dissertationen, aber auch über neue Forschungsbereiche und -themen diskutiert wurde.

Wir freuen uns als Conference Chairs, dass wir insgesamt 765 Teilnehmende in Siegen begrüßen durften, die das Event zu einer großartigen Veranstaltung gemacht haben. Aus den insgesamt 342 Full Paper Submissions wurden am Ende 119 Paper (Akzeptanzrate 34 %) ausgewählt und von den 81 Research-in-Progress Submissions schafften es am Ende 27 (Akzeptanzrate 33 %). Gleichzeitig wurden 25 Poster und 12 Demonstratoren akzeptiert, welche das Programm in zwei interaktiven Sessions ergänzten.

Die Gestaltung des Programms wäre ohne die 42 Track Chairs, die 272 Associate Editors und insgesamt 755 Reviewer nicht möglich gewesen. An dieser Stelle möchten wir uns herzlich für die tatkräftige Unterstützung bedanken. Der vorliegende Tagungsband fasst die Namen aller beteiligten Personen im Begutachtungs- und Auswahlprozess noch einmal zusammen.

Aus den Vorträgen heraus ergab sich eine Vielfalt an Themen, die auch auf den Social Events wie dem zum zweiten Mal stattfindenden Academic Networking Dinner und der erstmals im Format einer Empfangsgala veranstalteten Konferenzdinner mitsamt Poster und Demos zu einem regen Austausch führten.

Unser Dank gilt außerdem den zahlreichen Sponsoren und Freunden der Konferenz, welche durch ihre finanzielle Unterstützung die Konferenz überhaupt erst ermöglichten. Zusätzlich bedanken wir uns bei dem gesamten Organisationsteam, das über mehrere Monate vollen Fokus auf eine erfolgreiche WI 2019 gelegt hatten.

Abschließend bedanken wir uns bei allen Teilnehmenden, die über insgesamt fünf Tage die Wirtschaftsinformatik-Konferenz in Siegen zu einer tollen Veranstaltung gemacht und mit ihrer guten Stimmung die Abendveranstaltungen sowie das Rahmenprogramm Siegen bereichert haben.

Ihre Conference Chairs,

Thomas Ludwig & Volkmar Pipek

WI 2019 Team

Jun.-Prof. Dr. Thomas Ludwig

Conference Chair/ Sponsoring Chair

Prof. Dr. Volkmar Pipek

Conference Co-Chair

Prof. Dr. Gunnar Stevens

Program Chair

Victoria Wenzelmann

Workshop Chair

Maximilian Krüger

Demo & Prototype Chair

Michael Ahmadi

Poster & Research in Progress Chair

Prof. Dr. Martin Matzner
Emanuel Marx

Doctoral Consortium Chairs

Christoph Kotthaus

IT-Management Chair

David Unbehaun

Student Volunteer Chair

Konstantin Aal

Social Media Chair

Marc Gerbracht

Proceedings Chair

Dr. Daryoush Vaziri

Review Chair

Guido Kaufmann

Financial and Organizational Chair

Sven Hoffmann

Registration Chair

Merle Berg

Local Arrangements Chair

Qinyu Li

Design Chair

Kathrin Hoffmann

Publicity Chair

Nathanael Klein
Nico Castelli

Mobile App Chairs

Friederike Breuer

Communication Chair

Miriam Dell

Social Program Chair

Robin Arenz

Pub Crawl Chair

Saba Quehl

Operations Chair

Sponsoren der WI 2019

Track Chairs

Alexander Benlian | Benedikt Berger | Daniel Beverungen | Alexander Boden | Peter Burggräf | Paul Drews | Peter Fettke | Christoph M. Flath | Andreas Gadatsch | Natalia Kliewer | Hanna Krasnova | Dominik Kreß | Max Krüger | Sven Laumer | Christiane Lehrer | Thomas Ludwig | Alexander Mädche | Christian Maier | Ronald Maier | Emanuel Marx | Jorge Marx Gómez | Martin Matzner | Claudia Müller | Mario Nadj | Alena Otto | Erwin Pesch | Michael Prilla | Christian Reuter | Alexander Richter | Rene Riedl | Bernd Schenk | Sebastian Schlauderer | Andreas Schmidt | Daniel Schnurr | Marén Schorch | Matthias Schumann | Stefan Stieglitz | Ali Sunyaev | Matthias Trier | Nils Urbach | Daniel Veit | Jan vom Brocke | Markus Weinmann | Waldemar Zgrzebski

Associate Editors

Marc Adam | Frederik Ahlemann | Stephan Aier | Can Adam Albayrak | Paul Alpar | Rainer Alt | Wieland Appelfeller | Hans-Knud Arndt | H. Baars | Andrea Back | Christine Bauer | Annika Baumann | Sabine Baumann | Ulrike Baumöl | Jörg Becker | Daniel Beimborn | Markus Bick | Eva Bittner | Tilo Böhmann | T. Brandt | Stefan Brassel | Sebastian Bräuer | Christoph Breidbach | Michael H. Breitner | Rüdiger Breitschwerdt | Jörg Bremer | Jan Brinker | Christoph Buck | Nico Castelli | Annkristin Cordes | Tobias Dehling | Patrick Delfmann | Christoph Deutskens | Jens Dibbern | Mateucz Dolata | Andreas Drechsler | Paul Drews | Philipp Ebel | Katharina Ebner | Andreas Eckhardt | Mathias Eggert | J. F. Ehmke | Christian Ehnis | Christian Erfurth | Björn Escofier | Werner Esswein | Helmut Faasch | Benjamin Fabian | Michael Fellmann | Frank Fiedrich | Kathrin Figl | Hans-Georg Fill | A. Fink | Kai Fischbach | Thomas Fischer | Ulrich Frank | Detlev Frick | Gilbert Fridgen | Nora Fteimi | Frank Fuchs-Kittowski | Burkhardt Funk | Daniel Fürstenau | Stefan Geisler | Martin Gersch | Heiko Gewald | Henner Gimpel | Rob Gleasure | Jens Grossklags | Marcin Grzegorzek | D. Gutt | Steffi Haag | Janine Hacker | Jantje Halberstadt | Maik Hammerschmidt | Andre Hanelt | Sohaib Hassan | I. Heimbach | Heiner Heimes | Bernd Hellingrath | Dominik Herrmann | Georg Herzwurm | Thomas Hess | A. Hilbert | Sara Hofmann | Tim Höhne | Thomas Huber | Ursula Hübner | Marco Jahn | Christian Janßen | Florian Johannsen | Jürgen Jung | Reinhard Jung | Nina Kahnwald | Dimitris Karagiannis | Marc-André Kaufhold | Karlheinz Kautz | Michael Klafft | Stefan Klein | Thomas Klenke | Mathias Klier | Ralf Knackstedt | Mathias Knoll | Michael Koch | Ayelt Komus | Agnes Koschmider | Antonia Köster | Tobias Kowatsch | Oliver Krancher | Johann Kranz | Felix Kruse | Thomas Kude | Arno Kühn | Dennis Kundisch | Christoph Kunze | Martin Kütz | Matthias Laschke | Christoph Lattemann | Ulrike Lechner | Franz Lehner | Christiane Lehrer | Susanne Leist | S. Lessmann | Olga Levina | Michael Leyer | Luigi Lo Iacono | Peter Loos | Tim Majchrzak | Martin Manns | Oliver Marschollek | Christian Matt |

Martin Matzner | Jan Mendling | Thomas Meneweger | Tilo Mentler | Milad Mirbabaie | Andreas Möller | L. Mönch | Jürgen Moormann | Stefan Morana | Oliver Müller | Sara Nasiri | Simon Nestler | Bernhard Nett | Dirk Neumann | Peter Niemeyer | Klaus North | Markus Nüttgens | Andreas Oberweis | Sarah Oeste-Reiß | Nadine Ostern | Sven Overhage | Rainer Paffrath | Christina Pakusch | Hergen Pargmann | Jan Pawlowski | Christoph Peters | Jens Pöppelbuß | Oliver Posegga | Jens Pottebaum | Kai Pousttchi | Barbara Rapp | Asarnusch Rashid | Hans Peter Rauer | Michael Rebstock | Ulrich Remus | Christoph Rensing | Stefanie Rinderle-Ma | Susanne Robra-Bissantz | Maximilian Röglinger | Thomas Rose | Michael Rosemann | Kristina Rosenthal | Steffi Rudel | Stefan Sackmann | Alexander Sandau | Kurt Sandkuhl | Gerhard Satzger | Mario Schaarschmidt | Hannes Schlieter | Andreas Schmidt | Hendrik Scholta | Eric Schoop | Dirk Schreiber | G. Schryen | Monika Schuhmacher | Ronny Schüritz | Reinhard Schütte | Martin Semmann | T. Setzer | Markus Siepermann | Elmar Sinz | Stefan Smolnik | Matthias Söllner | Andreas Solsbach | Christian Soost | Ulrich Stache | Christian Stary | Martin Stein | Dennis Steininger | Dirk Stelzer | Michael Stepping | Stefan Stieglitz | Steffen Stock | Alexander Stocker | Stefan Stöckler | Susanne Strahringer | Özge Subasi | Lars Táxen | Hilda Tellioglu | Timm Teubner | Frank Teuteberg | Stefan Thalmann | Ferdinand Thies | F. Thiesse | Frédéric Thiesse | Oliver Thomas | Verena Tiefenbeck | K. Tierney | Klaus Turowski | Daryoush Vaziri | Melanie Volkamer | Johannes Wagner | Heinz-Theo Wagner | Benjamin Wagner vom Berg | Barbara Weber | Martin Weibelzahl | Christof Weinhardt | Markus Weinmann | Tim Weitzel | Helena Wenninger | Otilia Werner-Kytölä | Dirk Werth | Mathias Weske | Christian Weske | Lauri Wessel | Markus Westner | Thomas Widjaja | Rainer Wieching | Martin Wiener | Manuel Wiesche | L. Wiewiorra | Till Winkler | Volker Wohlgemuth | Jan Wolf | Felix Wortmann | Jochen Wulff | Stefan Wunderlich | Markus Zanker | Rüdiger Zarnekow | Steffen Zimmermann | Andreas Zolnowski

Reviewer

Puja Abbassi | Liliane Ableitner | Olga Abramova | Safa'a AbuJarour | Marc Adam | Martin Adam | Michael Adelmeyer | Michael Ahmadi | Suchit Ahuja | Stephan Aier | Ali Akyol | Simon Albrecht | Alexandro Alonso | Rainer Alt | Ireti Amojo | Jürgen Anke | Christian Anschütz | Gustavo Aragón | Hans-Knud Arndt | Jad Asswad | Andreas Auinger | Carolin Auschra | Peyman Badakshan | Florian Baer | Lucas Baier | Abayomi Baiyere | Mandy Balthasar | Wasana Bandara | Victoria Banken | Florian Bär | Benjamin Barann | Filipe Barata | Marcus Basalla | Olga Battaia | Christine Bauer | Katharina Baum | Annika Baumann | Frederik Bäumer | Tobias Baumgaertner | Hermann Baumgartl | Ulrike Baumöl | Jörg Becker | Felix Becker | Dennis Behrens | Daniel Beimborn | Mariem Ben Ben Rehouma | Benedict Bender | Dirk Bendlin | Martin Benedict | Ivo Benke | Alexander Benlian | Carina Benz | Peder Bergan | Matthias Berger | Cora Bergert | Lisa Berkemeier | Andrew Berry | Matthias Bertram | Matthias Betz | Jan Betzing | Benedikt Betzwieser | Markus Bick | Dominik Bilgeri | Matthias Billert | Nicola Bilstein | Eva Bittner | Kim Janine Blankenhagel | Michael Blaschke | Ivo Blohm | Maximilian Bock | Alina Bockshecker | Alexander Boden | Markus Böhm

| Tilo Böhmann | Lars Bollweg | Jens Borchers | Dominik Bork | Friedrich Born | Paul Bossauer | Sascha Bosse | Mohamed Bouras | Gerhard Brandhofer | Tobias Brandt | Daniel Braun | Tim Brauner | Jörg Bremer | Vincent Bremer | Sjaak Brinkkemper | Philipp Brune | Robert Buchmann | Erik Buchmann | Arne Buchwald | Christoph Buck | Tobias Buer | Hans-Ulrich Buhl | Peter Burggräf | Simone Burster | Peter André Busch | Sebastian Büttner | Ricardo Büttner | Dorothea Calmels | Buket Camurtay | Nico Castelli | Friedrich Chasin | Yi-Te Chiu | Sonja Christ-Brendemühl | Raffaele Ciriello | Catherine Cleophas | Ann-Kristin Cordes | Evgheni Croitor | Klaas Dählmann | Stelios Damalas | Matthias Dannapfel | Elena Davcheva | Triparna De Vreede | Maik Dehnert | Patrick Delfmann | Kevin Desouza | Christoph Deutskens | Stephan Diederich | Thomas Diefenbach | Soeren Diel | Barbara Dinter | Yvonne Dittrich | Mateusz Dolata | M. Dornhöfer | Mareike Dornhöfer | Cathal Doyle | Katharina Drechsler | Christian Dremel | Paul Drews | Christian Drumm | Jana Dücker | Alexander Dürr | Carolin Durst | Susanne Durst | Stefan Dyck | Aleksandra Dzepina | Philipp Ebel | Sebastian Eberz | Katharina Ebner | Martin Ebner | Mathias Eggert | Jan Fabian Ehmke | Christian Ehnis | Tom Eich | Mehdi Elahi | Andreas Emrich | Tobias Enders | Adrian Engelbrecht | Severin Engelmann | Christian Erfurth | Tatiana Ermakova | Sissy-Josefina Ernst | Werner Esswein | Helmut Faasch | Andri Färber | Filip Fatz | Masud Fazal-Baqaie | Jasper Feine | Carsen Felden | Carsten Felden | Niels Feldmann | Michael Fellmann | Helio Ferenhof | Marc Fernandes | Angela Fessl | Stefan Feuerriegel | Sven Feurer | Frank Fiedrich | Kathrin Figl | Hans-Georg Fill | Andreas Fink | Kai Fischbach | Lukas Fischer | Thomas Fischer | Lars Fischer | Marcus Fischer | Daniel Fischer | Simon Fischer | Christian Fitte | Ciara Fitzgerald | Christoph Flath | Malte Fliedner | Sebastian Floerecke | Jens Förderer | Matthias Förster | Ulrich Frank | Lena Frankenhauser | Daniel Franzmann | Detlev Frick | Michael Friedewald | Gregor Friedrich-Baasner | Albrecht Fritzsche | Jennifer Fromm | Michael Fruehwirth | Nora Fteimi | Davide Fucci | Matthias Fuchs | Frank Fuchs-Kittowski | Burkhardt Funk | Daniel Fürstenau | Andreas Gadatsch | Hannah Gaede | Bernhard Gahr | Kai Gand | Margaretha Gansterer | Sven Gehrke | Manfred Geiger | Stefan Geisler | Marc Gerbracht | Patrick Gerhards | Martin Gersch | Heiko Gewald | Nils Giesen | Henner Gimpel | Volkan Gizli | Margarita Gladkaya | Alexander Gleiß | Niklas Goby | Anat Goldstein | Valerie Graf | Tobias Greff | Anke Greif-Winzrieth | Matthias Griebel | Alexander Gröschel | Christian Grotherr | Marcin Grzegorzek | Nadine Guhr | Jens Gulden | Jana Gundlach | Sebastian Günther | Gunther Gust | Dominik Gutt | Steffi Haag | Knut Haase | Janine Hacker | Simon Hagen | Georg Hake | Jantje Halberstadt | André Halckenhäußer | Sebastian Halsbenning | Joshua Handali | Stefan Hanke | Jannis Hanke | David Harborth | Jörg Härtwig | Sohaib Hassan | Maximilian Haug | Matthias Hauser | Jennifer Hehn | Sven Heidenreich | Margareta Heidt | Heiner Heimes | Andreas Hein | Kai Heinrich | Lukas Hellwig | Felix Hemke | Susanne Hensely-Schinkinger | Thomas Hess | Manuel Hess | Markus Heuchert | Wilko Heuten | Andreas Hilbert | Matthias Hille | Knut Hinkelmann | Sebastian Hobert | Sven Hoffmann | Mario Hoffmann | Sara Hofmann | Tim Höhne | Peter Holz | Alexander Hölzemann | Heike Marita Hölzner | Valentin Holzwarth | Konstantin Hopf | Chris Horbel | Bettina Horlach | Lorenz Hörner | Olivia Hornung | Thomas Huber | Marvin Hubl | Alexander Hübner | Johannes Hübner | Veronika Huck-Fries | Philipp Hukal | Axel Hund | Thomas Hurni | Dominik Huth | Jan-paul Huttner | Sami Hyrynsalmi |

Fatih-Mehmet Inel | Andrew Isaak | Nihal Islam | Roope Jaakonmaeki | Axel Jacob | Stephan Jacobs | Katharina Jahn | Timo Jakobi | Christian Janßen | Florian Jasche | David Jaspert | Florian Johannsen | Jan Jöhnk | Marco Johns | Julia Jonas | Reinhard Jung | Jürgen Jung | Ekaterina Jussupow | Daniel Kaimann | Lena Kaiser | Lena Franziska Kaiser | Christian Kaiser | Matthias Kalverkamp | Michael Kamm | Dimitris Karagiannis | Marc-André Kaufhold | Alexander Keller | Rene Kessler | Sabine Khalil | Marjan Khobreh | Philipp Kienscherf | Thomas Kilian | Jasmin Kirchhübel | Tobias Kissmer | Michael Klafft | Adrian Klammer | Falco Klemm | Michael Klesel | Mathias Klier | Julia Klier | Simon Kloker | Benjamin Klör | Ralf Knackstedt | Bart Knijnenburg | Matthias Knoll | Kevin Koch | Jascha-Alexander Koch | Marc Kohler | Ayelt Komus | Jan Kopetz | Andrea Koppitz | Agnes Koschmider | Antonia Köster | Christoph Kotthaus | Nikita Kozodoi | Johann Kranz | Nane Kratzke | Mathias Kraus | Hannes Krause | Veronika Krauss | Dominik Kress | Robert Krimmer | Julia Kroenung | Tobias Kroll | Barbara Krumay | Jan-Peter Kucklick | Niklas Kühl | Stefan Kuhlemann | Babett Kühne | Oksana Kulyk | Dennis Kundisch | Sibylle Kunz | Christophe Kunze | Florian Künzler | Anna Kupfer | Christian Kurtz | Dilshod Kuryazov | Maria Kutar | Kati Kuusinen | Thomas Lampoltshammer | Claudio Lamprecht | Michael Lang | Fabian Lang | Joel Lanir | Isabel Laux | Gernot Lechner | Ulrike Lechner | Matthias Lederer | Nosan Lee | Stefanie Lehmann | Franz Lehner | Christiane Lehrer | Susanne Leist | Tim-Benjamin Lembcke | Christopher Lentzsch | Stefan Lessmann | Joerg Leukel | Olga Levina | Myriam Lewkowicz | Christian Leyh | Mahei Manhai Li | Jan-David Liebe | Kirsten Liere-Netheler | Georg Lindner | Luigi Lo Iacono | Peter Loos | Helena Lovasz-Bukvova | Patrick Lübbecke | Thomas Ludwig | Nick Lüthi | Hedda Lüttenberg | Bernhard Lutz | Thomas Lux | Stefan Mager | Ronald Maier | Christian Maier | Tim A. Majchrzak | Ahlam Malllak | Gabriela Mallmann | Martin Manns | Wybrands Marius | Nils Marscheider | Christoph Martin | Sabine Matook | Christian Matt | Jana Mattern | Dirk Christian Mattfeld | Florian Matthes | Stefan Mayer | Erika Meins | Jan Mendling | Tilo Mentler | Christian Meske | Tobias Mettler | Johanna Meurer | Kyrill Meyer | Corvin Meyer-Blankart | Bozena mierzejewska | Milad Mirbabaie | Tobias Mirsch | Mareike Möhlmann | Andreas Möller | Lars Mönch | Bernhard Moos | Stefan Morana | Dominik Morar | Pascal Moriggl | Tina Morlok | Lea Müller | Oliver Müller | David Müller | Jan Muntermann | Matthias Murawski | Patrick Murmann | Oliver Nalbach | Taewoo Nam | Sara Nasiri | Simon Nestler | Bernhard Nett | Dirk Neumann | Jürgen Neumann | Jörg Neumann | Tim Niesen | Fabian Nischak | Marcia Nißen | Mario Nolte | Markus Nuettgens | Indira Nurdiani Jabangwe | Irawan Nurhas | Niclas Nüske | Dumisani Nyumbeka | Andreas Oberweis | Caroline Oehlhorn | Caro Oehlhorn | Severin Oesterle | Nadine Ogonek | Daniel Olivotti | Christian Olt | Stefan Oppl | Frederike Marie Oschinsky | Lena Otto | Alena Otto | Boris Otto | Sven Overhage | Christina Pakusch | Malte Paskuda | Edoardo Patti | Tobias Pauli | Maria Pavlova | Alexandra Pehlken | Norman Pelzl | Nissrin Arbesun Perez | Erwin Pesch | Christoph Peters | Ralf Peters | Christian Peukert | Jella Pfeiffer | Nicolas Pfeuffer | Katharina Pflügner | Monika Pobiruchin | Jens Poeppelbuss | Oliver Posegga | Mathis Poser | Jens Pottebaum | Key Pousttchi | Ferry Pramudianto | Martin Prause | Hauke Precht | Michael Prilla | Christopher Prinz | Daniel Proba | Nicolas Pröllochs | Martina Prox | Yu Pu | Nils Randrup | Barbara Rapp | Rocco Raso | Daniel Rau | Julia Rauscher | Michael Rebstock

| Beke Redlich | Mobashar Rehman | Olaf Reinhold | Kathrin Reinke | Philipp Reinke | Andreas Reitz | Madeleine Renyi | Alexandra Rese | Mareike Reus | Melanie Reuter | Rainer Review | Renate Review | Robert Review | Alexander Richter | Peggy Richter | René Riedl | Dennis Riehle | Tim Rietz | Stefanie Rinderle-Ma | Malena Ripken | Marten Risius | Raphael Rissler | Robert Rockmann | Christoph Rohde | Thomas Rose | Michael Rosemann | Manfred Rosenberger | Kristina Rosenthal | Ewa Roszkowska | Angela Roth | Hannes Rothe | Steffi Rudel | Dominik Rüegger | Stefan Sackmann | Aditya Saharia | David Saive | Alexander Sandau | Kurt Sandkuhl | Christina Sarigianni | Gerhard Satzger | Mario Schaarschmidt | Harald Schallner | Johannes Schering | Raphael Schilling | Rolf Schillinger | Nina-Birte Schirrmacher | Hannes Schlieter | Andreas Schmidt | Günter Schmidt | Lisa Schmidthuber | Sabrina Schneider | Michael Schneider | Christoph Schneider | Johannes Schneider | Sofia Schöbel | Detlef Schoder | Hendrik Scholta | Michael Scholz | Paul Schott | Maximilian Schreieck | Melanie Schreiner | Monika Schuhmacher | Thomas Schulz | Michael Schulz | Ronny Schüritz | Reinhard Schütte | Florian Schwade | Stefan Seebacher | Isabella Seeber | Michael Seifert | Aurica Seiferth | Nikolaus Seitz | André Sekulla | Martin Semmann | Thorsten Semrau | Thomas Setzer | Sabine Seufert | Layal Shammas | Dominik Siemon | Markus Siepermann | Miroslav Sili | Mario Silic | Elmar Sinz | Martin Smits | Matthias Söllner | Andreas Solsbach | Ramanathan Somasundaram | Christian Soost | Alina Sorgner | Gabriele Sottocornola | Patrick Spieth | Michael Spitzer | Benjamin Spottke | Ulrich Stache | Philipp Staudt | Thorben Steenmann | Pia Steenweg | Lars Stegemann | Gerald Stei | Susanne Steigler | Martin Stein | Nikolai Stein | Dennis Steininger | Malgorzata Sterna | Tobias Steudner | Stefan Stieglitz | Mark Stieninger | Steffen Stock | Stefan Stöckler | Emanuel Stoeckli | Piotr Stolarski | Martin Stolk | Susanne Strahringer | Tim Straub | Jens Strüker | Gian Antonio Susto | Lukasz Swiatczak | Alfred Taudes | Sebastian Taugerbeck | Wissam Tawileh | Timm Teubner | Frank Teuteberg | Scott Thiebes | Tiemo Thiess | Christof Thim | Oliver Thomas | Tristan Thordsen | Matthias Tietz | David Tilson | Ingo J. Timm | Felix Timm | Peter Tinschert | Stefan Tönnissen | Stephen Treacy | Manuel Trenz | Constanze Tschöpe | Maria Tulis-Oswald | Klaus Turowski | Falk Uebernickel | Martin Ullrich | Marlin Ulmer | Soja Utz | Michel van der Borgh | Daryoush Vaziri | Georg Vetter | Nico Vitt | Pascal Vogel | Jannis Vogel | Kristin Vogelsang | Maximilian von Welck | Marta Vos | Franziska Voss | Michael Vössing | Johannes Wagner | Heinz-Theo Wagner | Gerit Wagner | Amina Wagner | Benjamin Wagner vom Berg | Lena Waizenegger | Daniela Waldmann | Sören Wallbach | Gianfranco Walsh | Thiemo Wambsganss | Kai Wang | Gaute Wangen | Jonas Wanner | Christian Weber | Kristin Weber | Peter Weber | Lars Wederhake | Charlotte Wehking | Benjamin Wehner | Martin Weibelzahl | Johannes Weibl | Julian Weidinger | Jake Weidman | Christoph Weinert | Janina Weingarth | Markus Weinmann | Peter Weiß | Katinka Weissenfeld | Jörg Weking | Karl Werder | Alexander Werner | Sebastian Werning | Oliver Werth | Mathias Weske | Lauri Wessel | Michael Wessel | Nora Wessels | Werner Wetzlinger | Eoin Whelan | Arnold Wibmer | Anna Wiedemann | Rouven Wiegard | Manuel Wiesche | Katharina Wiesneth | Sven Willrich | Axel Winkelmann | Rainer Winkler | Nane Winkler | Till Winkler | Andreas Winter | Jakob Wirth | Matthias Wissotzki | Anne-Katrin Witte | Volker Wohlgemuth | Isabell Wohlgenannt | Lena Wolbeck | Jan Wolf | Verena Wolf | Clemens Wolff | Jochen Wulf

| Philipp Wunderlich | Emrah Yasasin | Patryk Zapadka | Alex Zarifis | Rüdiger Zarnekow | Liudmila Zavolokina | Olaf Zawacki-Richter | Robert Zepic | Ruilin Zhu | Malgorzata Zieba | Oliver Zierke | Nicole Zigan | Alfred Zimmermann | Steffen Zimmermann | Elizaveta Zinovyeva | Andreas Zolnowski | Patrick Zschech | Oliver Zuchowski | Matthäus Zylka

Student Volunteers

Yasir Aurangzaib Abbasi | Michael Asante-Boateng | Joshua Blum | Katka Cerna | Petra Cernohorska | Martin Dickel | Tanja Ertl | Roxana Fischer | Jiannis Giatagantzidis | Julian Hester | Florian Jasche | Saskia Knipp | Jennifer Knött | Amanda Langer | Mariam Lehmann | Helia Nematollahi | Dominik Otto | André Sekulla | David Struzek | Sebastian Taugerbeck | Phi van Kieu ang Can | Majella Vincent | Nico Vitt | Anna Vogel | Tong Yu

Best Paper Award

Resolving Inconsistencies in Declarative Process Models based on Culpability Measurement
Carl Corea, Matthias Deisen, Patrick Delfmann

Who is Stressed by Using ICTs? A Qualitative Comparison Analysis with the Big Five Personality Traits to Understand Technostress
Katharina Pflügner, Jens Mattke, Christian Maier

Best Prototype/Demo Award

Privacy-friendly User Location tracking with Smart Devices: The Beat Prototype
Jan H. Betzing, Marco Niemann, Ingo Berendes

Best Poster Award

Towards Predictive Energy Management in Information Systems: A Research Proposal
David Heim, Gregor Friedrich-Baasner, Anna Fuchs, Axel Winkelmann

Best Student Paper Award

From Facets to Universality-An Approach to Defining IoT
Clarissa Endress

How Software Promotes the Integration of Sustainability in Business Process Management
Maren Stadtländer

Outstanding Associate Editor

Marc Adam, Paul Drews, Andre Hanelt, Heinz-Theo Wagner, Till Winkler

Outstanding Reviewer

Jörg Becker, Ann-Christin Cordes, Yi-Te Chiu, Jan Jöhnk, Raphael Rissler, Peter Tinschert

Inhaltsverzeichnis

Workshops **1990**

Track 1:

Produktion & Cyber-Physische Systeme

Track Chairs:

Prof. Dr. Thomas Ludwig
Universität Siegen

Prof. Dr. Peter Burggräf
Universität Siegen

Requirements and a Meta Model for Exchanging Additive Manufacturing Capacities

Chiara Freichel[1], Adrian Hofmann[1], Marcus Fischer[1], and Axel Winkelmann[1]

[1] Julius-Maximilians-Universität Würzburg, Lehrstuhl für BWL und Wirtschaftsinformatik, Würzburg, Germany
{chiara.freichel,adrian.hofmann,marcus.fischer,axel.winkelmann}@uni-wuerzburg.de

Abstract. In an environment shaped by digital transformation and globalization, manufacturers face increasing market dynamics, cost pressure, and more sophisticated customer requirements. As this demands flexibility and adaptability, enterprises rely on new solutions for collaboration. A marketplace for production capacities supports companies in reducing order risks and improving responsiveness to changing market conditions. We seek to define requirements for a marketplace that is capable of matching products with production processes. With an initial focus on additive manufacturing, we aim to build a blueprint for similar application scenarios in other industrial contexts. Therefore, we employ a qualitative research based on expert interviews. Our results suggest that a marketplace for production capacities must address various requirements, which can be grouped under the categories of technologies, machines, and products. We further build a conceptual meta model that sets the groundwork for the matching and thus facilitates the implementation of the marketplace in practice.

Keywords: Additive Manufacturing, Market Engineering, Sharing Economy, Sharing Capacities, Two-sided Markets

1 Introduction

In a globalized and digitalized business environment, manufacturers face various challenges that question established strategies and business models. At the center of their business activities, the area of production is influenced by many of the current megatrends, including *Industry 4.0* or *Internet of Things*. They do not only alter production concepts. Enterprises also face more sophisticated customer requirements, increasing competitive pressure, and higher market dynamics. The ability of manufacturers to respond to these changes is mainly determined by their production capacities. On the one hand, the number of orders can exceed available capacities, yielding delays or the non-fulfillment of orders with negative consequences on sales and profits. On the other hand, underutilization imposes the risk of fixed costs that cannot be fully covered. Overcoming the tradeoff between availability and production capacity related costs is frequently perceived as a key factor for the competitiveness of

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

manufacturers [1]. Furthermore, specialization has led to the emergence of supply networks, in which enterprises focus on their core competencies to contribute to and jointly build a complex product or service [2]. This comes with a division of tasks, activities, and responsibilities, which yields direct and indirect interdependencies between market participants and the emergence of single points of failure [3].

To cope with these challenges, enterprises rely on solutions that enable them to adapt to dynamic market conditions by offering or sourcing production capacities to other market participants. This does not only provide enterprises with the opportunity to align capacities with demands, but also fosters collaboration, communication, and the coordination of distributed actors.

Nowadays, complex material flows are supported and controlled by integrated information systems [1]. Most Industry 4.0 scenarios require a seamless integration of processes and information flows based on advanced information and communication technology. Thereby, production planning systems provide enterprises with an integrated database that contains information about available capacities, current delivery times, and future demand [4]. While these systems offer various functionalities for the company-specific management of production capacities, they hardly support transactions between multiple participants that collaborate in a supply network [1].

Against this backdrop, this study[1] seeks to conceptualize a marketplace for production capacities that operationalizes data from multiple manufacturers and yield improvements in various dimensions, such as performance, quality, and costs. Implemented as a central platform, the solution connects to enterprise software of different supply chain participants and allows for superior communication and collaboration. Based on the principles of the *Sharing Economy*, the proposed marketplace matches products with production processes as well as supply and demand. Similar solutions, such as Uber for transportation and Airbnb for housing, have proven effective in various contexts and application scenarios [5, 6]. Due to the complexity of the underlying research problem, we initially focus on the production technology of additive manufacturing. Additive manufacturing is more flexible and subject to less constraints. Hence, it provides the optimal starting point for this research endeavor. It allows us to build a proof of concept and to establish a conceptual foundation for similar solutions that address the requirements of other industries. Our research is guided by the following research questions (RQ):

1. What are the requirements for a marketplace that enables manufacturers to exchange production capacities in the field of additive manufacturing?
2. How can we design a meta model that matches products to production processes and thus facilitate the practical of implementation of a marketplace for production capacities?

---

[1] The study is based on experiences of the research project "DiHP – Dienstleistung für den integrierten Handel mit Produktionskapazitäten" (reference code 02K16C100). It is supported by the BMBF's programme "Technikbasierte Dienstleistungssysteme". Duration: 08/17-08/20.

To answer these RQ, we structure this study as follows: we present related work on platforms and collaborations in supply networks in the subsequent section. We describe the applied research design in Section 3 and present the results of our interview study in Section 4. Subsequently, we integrate our findings and build a meta model for the matching of products and production processes in Section 5. Ultimately, Section 6 concludes this study with a summary of findings, limitations, and future research potentials.

2 Foundations and Related Work

The concept of *Collaboration in Supply Networks* describes networking manufacturers, which has gained tremendous importance in recent years [7]. More sophisticated customer needs, considerable product development costs, and market dynamics yield an increasing pressure on supply networks and require them to realize continuous performance improvements [8]. Effective collaboration can leverage economies of scale, specialization, and integration [9–11]. The ongoing trend of the *Sharing Economy* has already led to a reorganization of a variety of established business activities [12]. Innovative technologies and business models incorporate the principles of this emerging paradigm and change industries by reducing transactions costs and providing easier ways to interact with customers. Thereby, products are increasingly augmented with services, which can be offered and sold based on various business models.

Electronic Commerce (e-commerce) is frequently described as a subclass of *Electronic Business (e-business).* E-business does not only cover the buying and selling of products and services, but also includes internet-based activities, such as customer service or electronic transactions [13]. E-commerce belongs to the e-business paradigm and describes the generation of revenues by initiating, negotiating, or processing business transactions via networks [14]. In 2015, e-commerce revenues amounted to 1,55 trillion US dollars and are currently expected to increase to 3,4 trillion US dollars in 2019 [15–17].

Collaborative Business (c-business) augments e-business or e-commerce with the concept of collaboration. It provides a holistic view on processes and inter-organizational cooperation. The concept uses the notion of end-to-end processes to integrate all participants of a supply network that collaborate and interact via the Internet, EDI, e-portals, or e-marketplaces. Represented by delivery deadlines or demand forecasts, collaborations can yield benefits in various dimension through the exchange of information [18].

Electronic Platforms in e-business can be divided into three types: the concept of *e-procurement* focuses exclusively on electronic purchasing, while *e-shops* pertain to sales activities. Furthermore, *e-marketplaces* provide platforms for the exchange of products and services. Products or services should be offered or requested without charge, but in an organized manner [14]. As of now, there is no marketplace that connects with established enterprise software and allows enterprises to exchange production capacities in order to address demand or supply fluctuations.

Building an e-commerce platform for exchanging additive manufacturing capacities requires profound knowledge about corresponding production processes, configurations, and influence factors. In general, manufacturing defines an industrial or artisanal production process and describes the way in which something is made. The production of components is specified in geometric and material terms [19]. *Additive Manufacturing*, commonly known as 3D printing, is one way of manufacturing. However, additive manufacturing summarizes a variety of manufacturing processes with the aim of producing physical objects on the basis of digital models that are converted into layer models by so-called *slicing*. Additive manufacturing processes add material in layers to obtain a desired object [20–23]. In the past, several techniques have been introduced that differ in regard to the way in which layers are created and connected. Because manufacturing works process oriented and depends upon the used material and machines, the dimensions machines, processes and materials are characterized by significant interdependencies [24].

3 Research Methodology

We employ a qualitative research design based on expert interviews to determine requirements for a marketplace that exchanges additive manufacturing capacities. The chosen design is a result of the limited knowledge on marketplaces for capacity sharing. Hence, we rely on an explorative research approach that integrates the perceptions and opinions of multiple stakeholders with extensive knowledge and experience in the domain of interest. The following section summarizes our research method and elaborates on the selection of interview partners, the design of the semi-structured questionnaires, and the procedures performed for data collection and analysis.

3.1 Selection of Interview Partners

Out of all interview partners four groups of companies and associated experts are defined and interviewed: (1) *potential buyers* of additive manufacturing capacities, (2) *potential sellers* of additive manufacturing capacities, (3) *producers of machines* for additive manufacturing and (4) *platform providers*. The interviewees are required to have knowledge about essential IT and management concepts as well as about the additive manufacturing technology. To select the experts, we first interview various employees of the nine companies to ensure they possess the necessary knowhow.

We interview potential buyers and sellers that already use additive manufacturing, intend to use it in the future, or are interested in buying additive manufacturing capacities. We integrate producers of additive manufacturing machines to evaluate aspects that are relevant for the underlying production processes and platform providers to shed light on requirements that must be addressed when implementing the proposed marketplace in practice.

We use multiple criteria, such as market position, experience in the field of additive manufacturing and e-commerce platforms, and industry and company size for case selection. This leads to a final sample of nine companies with one expert each. It consists of three companies from each of the two groups of potential buyers (buyer A, buyer B, buyer C) and sellers (seller A, seller B, seller C), two manufacturers (manufacturer A, manufacturer B), and one platform operator (platform operator A). There is only one platform operator existing on the market, which can serve as a basis and orientation for our research project. Users of the platform are the key actors on the platform and therefore reflect the most important group for the development of requirements. For this reason, three experts are interviewed there. Machine manufacturers all have similar structures and do not differ as much as the companies using the platform. For this reason, fewer experts are surveyed than in the category of users.

3.2 Questionnaire Design

The interviews are conducted by phone and are based on group-specific semi-structured questionnaires. The questionnaires contain basic information for the interviewees as well as guidelines and questions.

In order to ensure high quality and rigorousness, we operationalize the criteria proposed by [25]. Not included are the criteria *scaling*, *standardization*, and *fairness*, as they are primarily relevant for psychological analysis.

The following section describes the structure of the questionnaires. They are designed in a semi-structured manner, as the interviewees have a basic knowledge of the topics of interest. The interviews are conducted flexibly on the basis of open questions and provide the opportunity to include emerging concepts and ideas [26, 27]. In addition, the questionnaire contains different question types. Besides essential, factual, or direct questions, which address the key topic of this study, so-called *throw-away questions*, introductory, and structuring questions are used to guide the interview progress [28].

Each of the four questionnaires is introduced with a short preamble explaining the goals and scope of this study. In addition to a verbal explanation of the research, we provide information about the recording procedures, anonymity agreements, and data privacy and ask for the interviewees' approval. This is followed by two questions about their educational and work background. Thereby, we ask them to provide general information about their company and their function and responsibilities within it. This allows us to draw conclusions about the quality of their answers and to consider them in subsequent phases of this research.

The second section contains two general questions that introduce the topic and prepare the interviewees for more complex questions. Hence, we ask the interviewees about potential benefits of exchanging additive manufacturing capacities on the one hand, and about possible challenges on the other hand. These questions only serve the flow of discussion and the answers are used to justify this research but are not presented in more detail.

The subsequent part of the questionnaire varies with the expert group. Section three of the questionnaire for manufacturers and platform operators covers the portfolio of additive manufacturing technologies and machines. Questions about selected additive manufacturing methods, models, and production machines as well as about their characteristics are to be answered. In contrast, the third part for platform users includes questions about the buying and selling of capacities. Questions in this section analyze the external and/or internal additive manufacturing capacities used so far as well as their reasons for using a platform in this context.

The fourth section of the questionnaires for manufacturers and platform operators requests information about their machines, such as target groups and industries, applications, and covered product categories. Platform users are further asked about required processes and for characteristics that can be used to distinguish processes and machines.

The fifth part of the questionnaire for platform operators contains questions about the matching of products and additive manufacturing processes. The experts are asked for necessary/unnecessary attributes that must be considered when designing a matching mechanism. The questionnaire for manufacturers does not have a fifth section. Platform users are asked to provide information about the characteristics of additively manufactured products, including their material or size and other aspects.

3.3 Data Collection and Analysis

In order to achieve high-quality results, we describe the employed procedures for data collection and analysis subsequently [29]. Prior to the actual data collection, a pilot study was conducted with two independent researchers. The test confirmed that the questionnaire is understandable, well-structured, and of reasonable length and complexity [28]. As a result, we did not implement any significant changes to the design of the questionnaire.

Despite the various advantages of personal interviews as pointed out by [28], long distances required us to conduct the interviews via phone. The full conversations were documented based on audio recordings.

For data analysis, we followed the process explained in [30]. First, the available audio recordings were transcribed. The transcripts were written literally, the sentence structure was adapted, and filling words were deleted. Finally, we anonymized the transcripts.

The transcripts were then analyzed in order to derive requirements for the matching of products and production processes. For this purpose, the process of content analysis according to [30] was initiated by coding the data. Codes allowed us to break down relevant information to keywords. The coding process was performed by using the MAXQDA Version 18.0.8 software. We initially used the frequencies with which certain words were mentioned in order to determine suitable codes. We sorted the resulting words, consolidated synonyms, and classified sub-codes in subsequent steps. We then formed summarizing categories to group codes with similar meanings. In the next iteration, we controlled whether and to what extent all relevant questions of the guideline were covered by the codes and categories. The results of the categorization provided the basis to derive requirements for matching and sharing additive manufacturing capacities.

4 Requirements for Capacity Exchange

Subsequently, we introduce the requirements for a marketplace for capacity sharing and thereby answer RQ1. In summary, data analysis yielded a total of 371 coded text passages that are relevant for this research and thus suitable for requirements engineering. The passages are grouped into three categories *(1) characteristics of additive manufacturing technologies and machines* (192 occurrences), *(2) characteristics of products to be additively manufactured* (149 occurrences), and *(3) matching products and additive manufacturing technologies* (30 occurrences).

The analysis is carried out by means of a category-based evaluation along the main topics summarized of these three categories [30]. We define requirements for the process of developing the matching and thus exchanging additive manufacturing capacities via an e-commerce platform.

Recent studies identified more than 200 machine types for additive manufacturing on the market [24]. As a basis for expert evaluation and discussion we performed in-depth analysis of literature and actual technical documentation available, which cannot be mapped-out in further detail within the limitations of this paper. However, the findings served as the necessary conceptual foundation of clustering the multitude of additive manufacturing technologies, machine types and suitable characteristics.

Although all manufacturing processes are based on the manufacturing steps already explained, they differ e.g. in terms of how the respective layers are manufactured and bonded and which material is processed.

First of all, it is necessary to examine which additive manufacturing processes can be used for which groups of material.

Requirement 1: Additive manufacturing processes for plastic materials must be evaluated for exchanging manufacturing capacities in a first step, as these are most frequently used in practice.

The first code in category *(1) characteristics of additive manufacturing technologies and machines* contains selected manufacturing technologies represented by subcodes. These were derived on the basis of word frequencies. According to the experts, additive manufacturing processes for plastics are most frequently used in an industrial context. The companies surveyed named the technologies Stereolithography (SLA), Binder Jetting/3D Printing (BJ/3DP), Selective Laser Sintering (SLS), Fused Deposition Modeling (FDM), Multi Jet Fusion/High Speed Sintering (MJF/HSS), Multi Jet Modeling/PolyJet (MJM/PJ) and Digital Light Processing (DLP). Experts have described these as standard technologies. The material group *plastics* represents the largest group of materials for additive manufacturing and is therefore the first category to be considered [24, 31]. Furthermore, the subcode *plastics* in category *(2) characteristics of products to be additively manufactured* implies that buyers A, B and C mainly manufacture plastic products. The focus on the material group *plastics* could thus be decided.

Requirement 2: Additive manufacturing processes for metal materials must be included in a second step due to a high relevance in industrial production.

Following the manufacturing processes for plastics, experts attribute high importance to Selective Laser Melting (SLM) as manufacturing technology for processing metals. Among the interviewees, manufacturer B described the advantages of the SLM process in particular detail: "On the one hand, there is a large variety of materials that can be processed. You can process almost all types of metallics that can be pulverized. On the other hand, the devices are relatively flexible in their settings for different parameters. […] Speed is another aspect. We have some of the fastest devices on the market […] and also have a high build-up rate. This is a very high value, which of course also leaves some potential to be exploited. […] A great advantage is knowing that the speed can be further increased and the quality can be maintained at the same time." This was assigned according to the material group *metals* by means of codes and should be included in further research work due to its relevance. A second requirement arises in relation to the choice of additive manufacturing processes for a matching.

Requirement 3: The characteristic *material* must be considered separately due to various dependencies and specifications.

A total of 45 text passages were coded with the code *material* in category *(1) characteristics of additive manufacturing technologies and machines*. Material is considered by interviewees to be one of the most important parameters of additive manufacturing technologies and machines, which is why it is considered separately. According to manufacturer B, parameter settings, such as device temperatures, depend on the specific material within one group (e.g. metal or plastic). Different material variants or different manufacturers of similar materials do not have to be differentiated at first. Almost all experts emphasize the relevance of quality factors in connection with materials, especially plastics. Material properties include impact resistance, stability, food safety, density and roughness. Furthermore, according to platform operator A, the intended use of the products depends on the selected material: "The difficulty is that it is unclear what certain products are intended to be used for, how durable they have to be, and whether one has to stick to certain specifications, which again depends on the machine or material." In terms of the code *quality*, strength, elongation at break, food safety, heat resistance or biodegradability in this work can be defined as material properties and not as product properties. *Density* was mentioned by manufacturer B and platform operator A as one property of products. This is also regarded as material property.

Requirement 4: The property *material* should be used to link devices to sales offers and purchase requests.

Platform operator A suggests an intense relationship between devices and materials ("processing"). Hence, we must account for them when matching, selling, and exchanging additive manufacturing capacities. The choice of additive manufacturing machines is largely determined by the choice of one or more materials. The property *material* must therefore be related to the purchase order as well as devices. One or more materials can be selected for exactly one purchase request. In return, exactly one sales offer with one or more materials can be offered in the area of sales of additive production capacities, since not all materials to be processed for devices necessarily have to be offered by vendors. Again, one or more materials can be processed by one or more devices. This also results in the following requirement.

Requirement 5: For selling production capacities, it is relevant to specify the materials available in the company.

In the interview with manufacturer B especially, we found that it is important to account for the availability of materials in the company that sells production capacities. This confirms to requirement 4.

Requirement 6: The attribute *quality* is considered in the context of the two characteristics material and layer thickness, since these depend significantly on the material, manufacturing technology, and intended use.

Experts consider quality as an important criterion, but it is dependent on the material, intended use, and technology. Sellers A and B as well as manufacturer B use the term *quality* in relation to materials and manufacturing methods. This has to be considered in the selection of the characteristics in further research work. Manufacturer B concludes correspondingly with regard to an achievable quality in additive manufacturing processes: "There are no welding seams [in additive manufacturing], so that the parts [...] have a higher performance. This makes it more resilient than if it had a welding seam". Buyers A and B use the term *quality* to describe the outer appearance of the product, e.g. its geometry or surface finish. Manufacturer A mentions it in connection with the strength of the product. Seller B also places quality in connection with the intended use of a product. Furthermore, the factor *quality* must be defined differently for different purposes. A prototype, for example, has different quality requirements (high, medium, low) compared to a functional component. Figure 1 illustrates these dependencies.

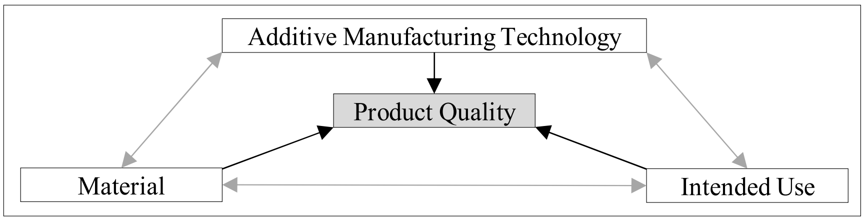


Figure 1. Factors Influencing Product Quality [32]

Within the scope of quality, sellers A, B and C consider it necessary to be aware of the purpose of a product or its category within the meaning of the subcode *product categories*. Often it is not possible to identify which products they manufacture. In this case, a request from the customer is necessary in order to be able to decide which production process is suitable for the respective application of the product in order to achieve the highest possible quality.

In relation to quality it is also necessary to investigate methods for post-processing, such as finishing, painting or assembly, which is represented by another code. In addition to platform operator A, the three groups of sellers named this characteristic. Seller B defines the removal or washing off of support structures in the FDM process as a standard process. Seller A, on the other hand, sees *CNC post-processing* as a necessary standard for increasing surface quality, which they could still not achieve. Seller C emphasizes that the material group *metals* in particular requires post-processing: "There are processes that are very new and more focused on the metal sector. This requires intensive mechanical post-processing, which we cannot do in-house, and therefore we have also made a certain selection there."

Requirement 7: Besides material group and material, multi-color, multi-material, and build volume dimensions/product size are relevant characteristics for matching additive manufacturing technologies and products.

As part of the code *multi-color* and *multi-material* in category *(1) characteristics of additive manufacturing technologies and machines*, the technology MJM/PJ offers the possibility of processing several colors during a production process, even if this is only in limited demand according to seller A. For seller B, the BJ/3DP method also enables multi-color. This confirms the statement that only a small number of procedures offer this functionality. Furthermore, a special feature in the area of plastics according to sellers A and C is the flexible material variety when using the MJM/PJ process. A distinction can be made between soft pressure and fixed pressure. According to the experts surveyed, most devices, with the exception of additional materials for support structures, have no functionality that supports processing more than one material. Platform operator A confirms the request for multi-material on the market. Seller A, on the other hand, experiences a rather low demand for multi-material.

In category *(2) characteristics of products to be additively manufactured* multi-color products are not relevant for buyers A, B and C. However, if this option is offered for a similar price, it would be an interesting option for the experts and thus, according to buyer C, a unique selling point of the products. The color is important for buyer B, as products are then painted, and the color of the base must therefore be accounted for. Similarly, buyer A emphasizes the importance offers that allow him to produce a desired color. This confirms the relevance of this criterion for future research.

Dimensions of the build volume serve as an important differentiator between various machines analyzed in category *(1) characteristics of additive manufacturing technologies and machines*. The space dimensions of the machines of all surveyed experts are smaller than one meter in height, width, and depth. The devices of manufacturer A, seller A, B and C can manufacture components of at least 5 mm in size. Sellers A defined typical sizes between 30 mm and 350 mm. These dimensions can be used as an average. A general statement on additive manufacturing technologies with predominantly larger or smaller dimensions cannot be formulated.

The experts agree that the size of the products for additive manufacturing is relevant for the selection process and can vary strongly. All experts of the group of buyers predominantly manufacture smaller products with approx. 5 mm to 200 mm in height, width and depth. Dimensions between 5 mm and 800 mm were mentioned. Platform operator A emphasized the fact that a product model which did not originally fit into the installation space can be made suitable by rotation or other slicing. In addition, a product that is too large could be manufactured in several parts and then be assembled.

Requirement 8: Production costs and time are one criterion for selecting additive production devices in comparison to conventional production processes and constitute an important requirement for the platform.

Another relevant code of category (1) *characteristics of additive manufacturing technologies and machines* is costs. Buyer B emphasized that by introducing the e-commerce platform, consideration would be given to the purchase of additive manufacturing devices. In addition, according to buyers B and C, the current production costs are one criterion for selecting additive production devices in comparison to conventional production processes. A distinction is also made between additive manufacturing processes with regard to acquisition costs as well as manufacturing and material costs. Manufacturer A describes the FDM process as the most cost-effective. Cost and price mechanisms on the developed platform are not considered in this work as a first analysis, but according to the interviewees they are an important factor, which will be included in further research.

Manufacturers A and B distinguish between machines according to their speed / production time, e.g. influenced by the size of the print head or the number and strength of the lasers in the SLM process. Buyer C considers speed to be an important factor in the purchase of additive manufacturing capacities. The relevance of the delivery time was emphasized according to buyer B: "For us only the delivery time of products is relevant. Of course, it depends on the priority of the order. [Normally we need our products] within two weeks." This was also emphasized in the sense of the code *production time*. Likewise, buyer C attaches importance to a certain speed. No distinction was made for buyer C between production and delivery time. Since production speed is dependent on the used technology, material, and quality in the sense of layer thicknesses and therefore require comprehensive analyses for comparability, separate research should be carried out in this respect.

Requirement 9: There is increasing potential in the application area of additive manufacturing for small series as well as series production. Therefore, the attribute quantity must be included in this analysis.

In a first step, the subcodes *prototypes*, *customized production*, *spare parts*, *small serie*s and *series production* are used to determine the purpose for which additively manufactured products are used. According to the experts, the products are mainly used as prototypes, individual production and spare parts. According to buyers A and C, platform operator A and seller C, there is increasing potential in the area of application of small series and series production. Manufacturer B has evolved from its original focus on prototype types to a manufacturer of additive production equipment for the production of functional components. Buyer B only focuses on individual production, which is in the focus of the company's business model. Both manufacturer A and seller A build spare parts or individual products for their own requirements for additive production machines. Small series and series production are often related to quantities. For example, platform operator A applies small series for quantities of 10 to 100 parts, while buyer A suggests that using additive manufacturing is beneficial even when producing series containing more than 100,000 parts.

Requirement 10: The platform should support application scenarios in different industries.

The subcode *industries* suggests that additive manufacturing is used across industries. The interviewees agree that the automotive industry especially benefits from the use of additive manufacturing. Buyers A, manufacturer B and platform operator A also name the medical, dental, and aviation industries as leading users of additive manufacturing.

5 Meta-Model

Based on the identified requirements, we can define a meta model that supports exchanging production capacities over an e-commerce marketplace. Thereby, *Sales Offers* reflect additive production machines or technologies and *Purchase Enquiries* reflect products. In order to ensure the structured and consistent storage of data and the efficient use of the database, this section introduces a conceptual data model shown and thereby answers RQ2.

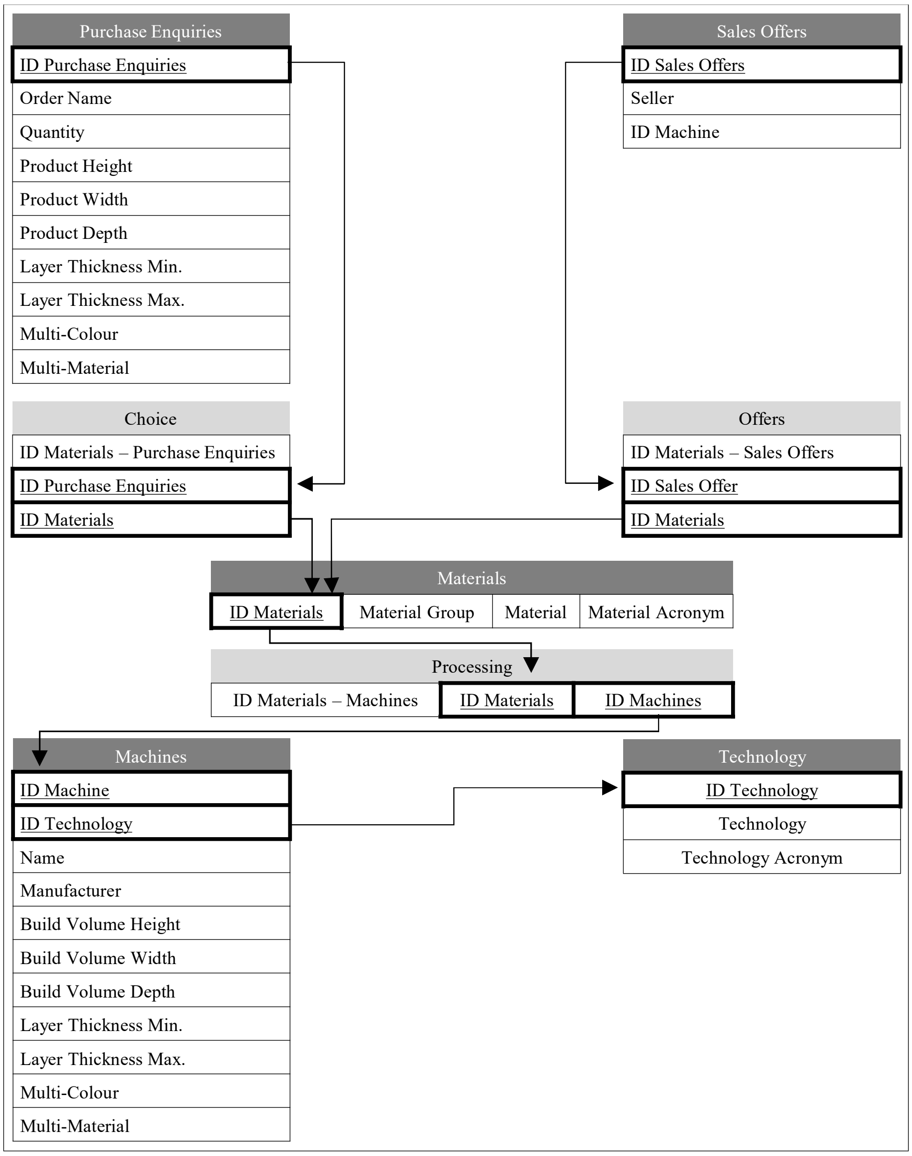


Figure 2. Meta-model for matching products and additive manufacturing technologies

Figure 2 summarizes the resulting model for matching machines with products in a structured collection of tables that are connected through relations. The characteristics of the *Purchase Enquiries* and *Sales Offers* were derived from the developed requirements. The choice of additive manufacturing *Machines* is mainly determined by the choice of *Materials*. These have therefore been added as a separate table and are directly related to *Machines, Purchase Enquiries* and *Sales Offers,* each connected via junction tables (*Processing*, *Choice*, *Offers*). *Materials* can be selected for a *Purchase Enquiry*. In return, a *Sales Offer* with one or more *Materials* can be offered in the area of capacity sales, because not all materials that are processed for devices necessarily have to be offered by vendors. Material properties have to be accounted for in future studies. Again, *Materials* can be processed by *Machines*. The *Machines* have the analyzed attributes and are assigned to *Technologies*.

6 Conclusion

A marketplace for production capacities can significantly contribute to the competitiveness of manufacturers and reduce risks in interwoven supply networks. It supports companies in accomplishing their strategies and business goals, such as a reduction of order risks caused by equipment failure or adaptability in case of changing market demands. Furthermore, they can benefit from synergies, economies of scale, and economies of scope.

This study set out to build the conceptual foundation for a marketplace solution for buyers and sellers of production capacities from the domain of additive manufacturing. Therefore, we sought to develop a catalog of requirements for a workable platform solution (RQ1) as well as to build a meta model that allows the matching of products and production processes in practice (RQ2). To answer RQ1, we conducted a qualitative study based on expert interviews. We questioned experts from nine companies that provided valuable insights from the perspectives of market participants, platform operators, and machine producers. As a result, we derived 10 requirements that can be grouped under the categories of *characteristics of additive manufacturing technologies and machines*, *characteristics of products to be additively manufactured*, as well as *matching products and additive manufacturing technologies*. A marketplace for production capacities must address these requirements to be adopted and to facilitate the expected benefits. To answer RQ2, we integrated our findings and built a meta model that supports the matching of products and production processes. It consists of 8 components and provides a conceptual foundation for the implementation of a matching algorithm in practice. By addressing a white spot in the current IS literature, this study represents a suitable starting point for future research endeavors that aim to examine requirements for exchanging production capacities in other scenarios and industry contexts.

This research is not without limitations. First, it is exploratory by nature and builds upon the perceptions and conclusions of the researcher. Other researchers may have derived different implications and requirements. Second, with only nine participants, our data sample is yet too small to derive a definite set of requirements for exchanging production capacities. Larger studies with more participants are necessary to provide definite evidence on the validity and reliability of our findings. Third, the resulting meta model is of preliminary nature and only provides an initial conceptualization of the marketplace. In future studies, we seek to implement, evaluate, and refine its properties. Future research opportunities lie primarily in transferring our findings to other industries and business scenarios. While its relevance is constantly growing, additive manufacturing only accounts for a small portion of industrial production capacities. As a consequence, future research must incorporate larger industries to realize the full potential of production capacity sharing.

References

1. Schuh, G., Potente, T., Reuter, C., Hauptvogel, A.: Steigerung der Kollaborationsproduktivität durch cyber-physische Systeme. In: Vogel-Heuser, B., Bauernhansl, T., und ten Hompel, M. (Hrsg.) Handbuch Industrie 4.0 Band 2. S. 75–92. Springer, Berlin, Heidelberg (2017).
2. Basole, R.C., Ghosh, S., Hora, M.S.: Supply Network Structure and Firm Performance: Evidence From the Electronics Industry. IEEE Transactions on Engineering Management. 65, 141–154 (2018).
3. Li, Y., Chang, Q., Xiao, G., Arinez, J.: Data-Driven Analysis of Downtime Impacts in Parallel Production Systems. IEEE Transactions on Automation Science and Engineering. 12, 1541–1547 (2015).
4. Chen, J.-H.: Simultaneous optimization of production planning and inspection planning for flexible manufacturing systems. In: Proceedings of the 9th annual conference on Genetic and evolutionary computation. S. 1928–1935. ACM Press, New York (2007).
5. Oskam, J., Boswijk, A.: Airbnb: the future of networked hospitality businesses. Journal of Tourism Futures. 2, 22–42 (2016).
6. Cusumano, M.A.: How Traditional Firms Must Compete in the Sharing Economy. Communications of the ACM. 58, 32–34 (2015).
7. Schmitt, R., Humphrey, S., Ellerich, M., Groggert, S.: Kapazitätsmarkt – Ressourcenhandel für die Produktion. Eine Cloud-basierte Plattform zum unternehmensübergreifenden Austausch von Produktionskapazitäten. Industrie 4.0 Management. 31, 30–34 (2015).
8. Grigoriev, A., Jung, V., Peeters, M., Vredeveld, T.: On the Acceptance of Gain Sharing Methods in Supply Chain Collaboration. GSBE Research Memoranda. (2017).
9. Logan, M.S.: Using Agency Theory to Design Successful Outsourcing Relationships. The International Journal of Logistics Management. 11, 21–32 (2000).
10. Soylu, A., Oruç, C., Turkay, M., Fujita, K., Asakura, T.: Synergy analysis of collaborative supply chain management in energy systems using multi-period MILP. European Journal of Operational Research. 174, 387–403 (2006).
11. Chen, L., Zhao, X., Tang, O., Price, L., Zhang, S., Zhu, W.: Supply chain collaboration for sustainability. A literature review and future research agenda. International Journal of Production Economics. 194, 73–87 (2017).
12. Stephany, A.: The Business of Sharing. Making it in the New Sharing Economy. Palgrave Macmillan, Basingstoke, England (2015).
13. Turban, E., Outland, J., King, D., Lee, J.K., Liang, T.-P., Turban, D.C.: Electronic Commerce 2018. Springer International, Cham, Switzerland (2018).

14. Heinemann, G.: Der neue Online-Handel. Geschäftsmodelle, Geschäftssysteme und Benchmarks im E-Commerce. Springer Fachmedien, Wiesbaden (2018).
15. Van Alstyne, M.W., Parker, G.G., Choudary, S.P.: Pipelines, Platforms, and the New Rules of Strategy. Harvard Business Review. April, 54–62 (2016).
16. Statista GmbH: Retail e-commerce sales worldwide from 2014 to 2020 (in billion U.S. dollars), https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/.
17. Lin, Z., Wang, Q.: E-commerce Product Networks, Word-of-mouth Convergence, and Product Sales. Journal of the Association for Information Systems. 19, 23–39 (2018).
18. Silberberger, H.: Collaborative Business und Web Services. Ein Managementleitfaden in Zeiten technologischen Wandels. Springer, Berlin (2003).
19. Fritz, A.H., Schulze, G.: Fertigungstechnik. Springer, Berlin; Heidelberg (2008).
20. Lasi, H., Morar, D., Kemper, H.: Additive Manufacturing. Herausforderungen für die gestaltungsorientierte Wirtschaftsinformatik. In: Tagungsband Multikonferenz Wirtschaftsinformatik 2014. S. 417–428. , Paderborn (2014).
21. Petrovic, V., Gonzalez, J.V.H., Ferrando, O.J., Gordillo, J.D., Puchades, J.R.B., Griñan, L.P.: Additive layered manufacturing. Sectors of industrial application shown through case studies. International Journal of Production Research. 49, 1061–1079 (2011).
22. Atzeni, E., Salmi, A.: Economics of additive manufacturing for end-usable metal parts. The International Journal of Advanced Manufacturing Technology. 62, 1147–1155 (2012).
23. Wirth, M.: Additive Fertigung. Technologie, Markt und Innovation. Julius-Maximilians-Universität Würzburg, Würzburg (2017).
24. Gebhardt, A., Kessler, J., Thurn, L.: 3D-Drucken. Grundlagen und Anwendungen des Additive Manufacturing (AM). Carl Hanser, München (2016).
25. Moosbrugger, H., Kelava, A.: Testtheorie und Fragebogenkonstruktion. Springer, Berlin, Heidelberg, New York (2012).
26. Paré, G.: Investigating Information Systems with Positivist Case Study Research. Communications of the Association for Information Systems. 13, 233–264 (2004).
27. Edwards, R., Holland, J.: What is Qualitative Interviewing? Bloomsbury Academic, London, New York (2013).
28. Berg, B.L.: Qualitative Research Methods for the Social Sciences. Allyn & Bacon, Needham Heights (2001).
29. Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. MIS Quarterly. 11, 369–386 (1987).
30. Kuckartz, U.: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Beltz Juventa, Weinheim, Basel (2014).
31. Statista GmbH: Meistgenutzte 3D-Druck-Technologien im Jahr 2017, https://de.statista.com/statistik/daten/studie/760408/umfrage/meistgenutzte-3d-druck-technologie/.
32. Witt, G.: Handlungsfelder. Additive Fertigungsverfahren, https://www.vdi.de/fileadmin/vdi_de/redakteur_dateien/gpl_dateien/6112_PUB_GPL_Thesen_und_Handlungsfelder_-_Additive_Fertigungsverfahren_Internet.pdf.

Service Systems, Smart Service Systems and Cyber-Physical Systems—What's the difference? Towards a Unified Terminology

Dominik Martin[1], Robin Hirt[1], and Niklas Kühl[1]

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Karlsruhe, Germany
{martin,hirt,kuehl}@kit.edu

Abstract. As businesses and their networks transform towards co-creation, several concepts describing the resulting systems emerge. During the past years, we can observe a rise of the concepts Service Systems, Smart Service Systems and Cyber-Physical Systems. However, distinct definitions are either very broad or contradict each other. As a result, several characteristics appear around these terms, which also miss distinct allocations and relationships to the underlying concepts. Previous research only describes these concepts and related characteristics in an isolated manner. Thus, we perform an inter-disciplinary structured literature review to relate and define the concepts of Service Systems, Smart Service Systems and Cyber-Physical Systems as well as related characteristics. This article can, therefore, serve as a basis for future research endeavors as it delivers a unified terminology.

Keywords: Service System, Smart Service System, Cyber-Physical System, literature review, conceptualization

1 Introduction

As businesses become interconnected, new opportunities and challenges arise for collaboration and co-creation [1, 2]. Different concepts, such as (Smart) Service Systems [3, 4] and Cyber-Physical Systems [5] emerge and strive to allocate, structure and explain phenomena in the field of digitally interconnected systems. However, these concepts are often used synonymously [4, 6] or contradict each other [5, 7]—which can lead to confusion and misunderstandings among practitioners and researchers. As a clear distinction of those concepts and related characteristics fosters the quality of future research, we aim to distinct Service Systems, Smart Service Systems and Cyber-Physical Systems. Thus, we ask the research question of *"How are the concepts Service System, Smart Service System and Cyber-Physical System defined and interrelated?"*. To approach this topic, we perform a structured literature research based on vom Brocke et al. [8] and Cooper [9] to identify commonly used definitions. We consolidate the insights and define each concept on this basis.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Additionally, we aim to derive a conceptualization including the three concepts and ask the additional question: *"Which characteristics are mentioned in the context of the concepts?"*. By applying an open coding approach [10] on 110 identified articles from different disciplines defining the concepts, we identify several characteristics that are mentioned in literature and allocate them accordingly.

We aim to provide distinct definitions of these concepts in order to set a foundation for researchers and practitioners to understand the terms consistently. Based on this, we intend to overcome boundaries to other disciplines and allow for a common understanding as well as, accordingly, to accelerate new research and development in these areas.

The remainder of this article is structured as follows: First, we present theoretical foundations regarding socio-technical systems and system-of-systems. Second, we describe our methodology comprising a literature search followed by an open coding analysis of the identified concept definitions. Third, we analyze all three concepts in isolation and then summarize them through a conceptualization. Fourth, we present a discussion followed by a conclusion.

2 Theoretical Foundations

A system is generally referred to as a "collection of components organized to accomplish a specific function or set of functions" [11, p. 73]. Boulding [12] particularly stresses the system boundaries which delimit a system and determine which parts belong to a system and which to the environment. In an open system, interactions can take place with the environment, whereas in an isolated system no interactions can take place [11]. Interactions can be both the exchange of information (from an Information Systems (IS) viewpoint) [11] and the exchange of mass or energy (from a nature science viewpoint) [13]. Particularly complex open systems consisting of multiple parts that perform complex interactions with each other and with the environment are widely spread in reality [14].

In order to categorize (Smart) Service Systems and Cyber-Physical Systems and form a better understanding of these terminologies, the basic concepts *socio-technical systems* and *system-of-systems* are introduced.

2.1 Socio-technical Systems

The term socio-technical system is often used to describe complex systems consisting of several interacting components [15]. Originally, however, the term was used to describe a set of people and related technologies that are structured in a certain way to produce a specific result [16].

According to Cartelli [17], a socio-technical system consists of two components (subsystems): The technical subsystem represents assets such as machines and equipment, as well as processes and tasks that are responsible for the conversion of input resources into outputs. The social subsystem is made up of people (such as employees) who are structured in groups and have assigned certain roles to operate, control and use the components of the technical subcomponent. Cartelli emphasizes the facet of knowledge, which is "socially constructed and developed in the interactions among people" [17, p. 3], as part of the social subsystem and its value for a socio-technical system.

Both subsystems are "jointly independent, but correlative interacting" [16, p. 17] in order to pursue and adapt to goals in the socio-technical system's environment and are therefore not separable from each other due to their manifold dependencies [15].

2.2 Systems-of-Systems

A system-of-systems has—like a typical system—interdependent components operating together to accomplish a certain common goal [18]. Unlike a typical system, the components of a system-of-systems are themselves systems [18]. According to Maier [19] a system-of-systems is an "assemblages of components that are themselves significantly complex, enough so that they may be regarded as systems and that are assembled into a larger system" [19, p. 269]. However, Maier names two limitations: First, the components must be operationally independent. That is, if a system-of-systems is broken down into its components, they must be able to fulfill their original purpose independently. Second, the component systems can not only work independently of each other, they do so as well. Thus, the subsystems maintain their operational independence continuously. Gideon et al. [18] summarize a system-of-systems as a "system build from independent systems that are managed separately from the larger system" [18, p. 357].

3 Methodology

With the foundations of socio-technical systems and systems-of-systems set, we elaborate on our applied methodology to reconstruct the state of the art of relevant literature. The scope of our literature review is systematized by the taxonomy proposed by vom Brocke et al. [8] and Cooper [9]. This taxonomy consists of six characteristics that distinguish literature reviews—focus, goal, organization, perspective, audience, and coverage—each including specific categories. Some of these categories are mutually exclusive, while for other characteristics several categories can be combined.

The focus of our literature research corresponds to the category research outcomes of the above-mentioned taxonomy. Furthermore, the goal of this article is the aggregation of already existing articles on the concepts Service System, Smart Service System and Cyber-Physical System—as well as their integration. The organization of this article is conceptually structured in order to aggregate the concepts separately. This article takes a neutral perspective. The target audience are scholars who are in need of a clear definition of the concepts as well as their distinction. To provide an appropriate overview of existing research, the literature search covers selected conferences and journals and, therefore, aims to be representative.

We conduct a systematic literature research according to vom Brocke et al. [8] in July 2018. While doing so, we focus on peer-reviewed articles from the field of

Information Systems, Service Science and Computer Science. In order to receive articles elaborating on (Smart) Service Systems and Cyber-Physical System, we use the search query: *"Service System" OR "Smart Service System" OR "Cyber Physical System"*. In a first step, we focus our search on the following selected Information Systems conferences and journals: International Conference on Information Systems (ICIS), European Conference on Information Systems (ECIS), Hawaii International Conference on System Sciences (HICSS), Information Systems Research (ISR), Information Systems Journal (ISJ), Management Information Systems Quarterly (MISQ), Journal of Management Information System (JMIS), European Journal of Information Systems (EJIS) and Business and Information Systems Engineering (BISE).

It is noticeable that the conferences have a much higher proportion of hits (ECIS: 21, ICIS: 14, HICSS: 10) in total than the journals (BISE: 8, MISQ: 2). The journals ISR, ISJ, JMIS and EJIS have no hits at all. In addition, it is recognizable that most of the hits date from the year 2018. Moreover, the number of hits has increased (Figure 1) over the years, which implies a strong relevance in terms of timeliness and strengthens the necessity for a clear nomenclature.

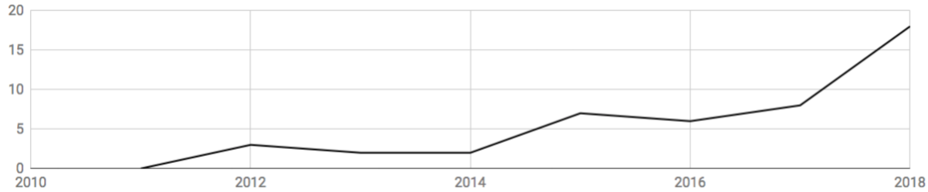


Figure 1. Number of hits in selected IS journals

When analyzing the outcomes, we noticed most of the articles relate to the concept of Service Systems, while in relatively few results the terms Smart Service System or Cyber-Physical System appear. Therefore, we extend our search across all disciplines using the literature database *Web of Science*. We realize that the term Service System plays a dominant role in the IS community, whereas the concept Cyber-Physical System occurs mainly in Computer Science literature. However, the term Smart Service System barely appears in the Web of Science database. Based on these findings, we conduct a Web of Science search for each of the three concepts separately and sort the results by number of citations and thereupon append the first 50 results for each concept to the literature list as well. In a third step, in addition, the outlets from the disciplines Service Science (six outlets with impact factor above 1[1]) and Computer Science (22 outlets with impact factor above 5[1]) are included as well. Thus, we ensure each of the communities in which the concepts are mainly used, are represented in this literature overview accordingly in a balanced manner.

Overall, the applied methodology results in an amount of 354 articles, which are selected by reading the abstract in order to exclude unrelated articles. Through forward

[1] The lower threshold for included outlet's impact factors is derived by the multiplication of the highest impact factor achieved in the specific discipline with a factor of 20 %

and backward search, further relevant articles are identified. By completely reading the remaining articles, all in all 110 relevant articles are selected and analyzed in a final step. All passages containing definitory statements with regard to at least one of the concepts is further analyzed by a coding approach to derive characteristics and interrelations of the considered concepts.

Within this subsequent step, two researchers analyze the concept definitions extracted from the articles using an open coding approach according to Saldaña [10]. With open coding we aim to find recurring characteristics of the individual concepts [10]. At the same time, we try to stay as open and unconstrained as possible in order to identify outstanding and particularly common characteristics from the literature. During this phase, we constantly compare all codes coded by two researchers as well as the underlying concept definitions to cluster passages that pertained to common codes. To substantiate our findings, we further integrate these common codes in order to derive more abstract conjoint categories and to harmonize different views.

4 Results

The results of the literature search and the analysis of the definitions depicted in each article are summarized in this section. In order to provide the reader with a comprehensive picture of the differences and similarities of the definitions, first the concepts are considered individually, before they are compared with each other.

4.1 Service Systems

The concept *Service System* appears most frequently in the results of our conducted literature search. Overall, 64 articles refer to the term Service System. According to Spohrer et al. [3] a Service System comprises *"service providers and service clients working together to coproduce value in complex value chains or networks"* [3, p. 72]. Components of a Service System are *"people, technology, internal and external service systems connected by value propositions, and shared information"* [3, p. 72] and examples include individuals, firms and nations. Based on this article from 2007, Maglio and Spohrer [20] synthesize the definition and formulate: *"Service systems are value-co-creation configurations of people, technology, value propositions connecting internal and external service systems, and shared information (e.g., language, laws, measures, and methods)"* [20, p. 18]. Examples include cities, businesses, nations, as well as individuals as the smallest representative of a service system and world economy as the largest [20].

The majority of articles adopt this definition [4, 7, 21–25], while others phrase it slightly different, but in principle remain faithful to the overall message [26–33]. Besides the more detailed definitions, some authors like Kleinschmidt and Peters [34] and Lintula et al. [35] use shorter and thus less specific descriptions. Böhmann et al. [36], Dörbecker and Böhmann [37] and Li and Peters [38] state that a Service System is a *"socio-technical system that enables value co-creation guided by a value* *proposition"* [36, p. 74], whereas Brust et al. [32] describe it as *"collections of people, technology and interactions"* [32, p. 8].

However, some authors deviate from this common definition and suggest divergent definitions, such as the one proposed in Höckmayr and Roth [39]: *"A service system is composed of multiple entities that interact to co-create value"* [39, p. 3]. Similarly, Motta et al. [40] differs from the common definition and describe a Service System only very abstract as a system which supports business services. Alter [41–44] refers to work systems and defines Service Systems as *"work systems that produce product/services and that may or may not involve co-production by customers and value co-creation"* [41, p. 4], while a work system is a *"system in which human participants and/or machines perform work using information, technology, and other resources to produce products and services for internal or external customers"* [41, p. 4]. Although some authors like Blohm et al. [45], Dörbecker et al. [46] and Matzner and Scholta [47] use the term Service System and name components as well as properties, but avoid defining it.

In conclusion, we also suggest using the definition according to Maglio and Spohrer [20] and Spohrer et al. [3], as it is the most concise and commonly used one, and define Service Systems for this article as *"value-co-creation configurations of people, technology, value propositions connecting internal and external service systems, and shared information (e.g., language, laws, measures, and methods)"* [20, p. 18].

4.2 Smart Service Systems

The concept *Smart Service System* has the lowest number of hits with only 10 represented articles in the searched outlets and databases. This concept is described by Barile and Polese [7], Maglio [4] and Medina-Borja [48] as an extension of the Service System concept containing self-management capabilities. Barile and Polese [7] define: *"Smart service systems may be intended as service systems designed for a wise and interacting management of their assets and goals, capable of self-reconfiguration (or at least of easy inducted re-configuration) in order to perform enduring behavior capable of satisfying all the involved participants in time"* [7, p. 31].

According to Maglio [4], Smart Service Systems are *"capable of self-detection, self-diagnostic, self-corrective, or self-controlled functions through the incorporation of technologies for sensing, actuation, coordination, communication, control, and more"* [4, p. 1]. By automating and self-managing systems, high costs and security risks caused by humans can be reduced, which can lead to improved offers or even new ones [4].

Beverungen et al. [49] state that Smart Service Systems are Service Systems, *"in which smart products are boundary-objects that integrate resources and activities of the involved actors for mutual benefit"* [49, p. 6].

According to the authors Maglio and Lim [50] as well as Medina-Borja [48], such a system is even *"capable of learning, dynamic adaptation, and decision making based upon data received, transmitted, and/or processed to improve its response to a future situation"* [50, p. 2], which can be done by integration of sensing, actuation and communication technologies. In addition, Maglio and Lim [50] describe that big data analytics can contribute to the innovation of Smart Service Systems by *“embedding human knowledge and capabilities in technologies to serve human purposes for effective value co-creation”* [50, p. 3]. Santo et al. [51] also emphasize the capability of such a system to learn and to *“simultaneously optimizing the use of resources and improving the quality of the services provided”* [51, p. 3].

Nevertheless, we recommend using a modification of the definition proposed by Medina-Borja [48] as it is the most detailed and comprehensive and includes most of the characteristics of the other definitions. Furthermore, it delivers a clear demarcation from Service Systems: *“A 'smart' service system is a [Service] [S]ystem capable of learning, dynamic adaptation, and decision making based upon data received, transmitted, and/or processed to improve its response to a future situation. The system does so through self-detection, self-diagnosing, self-correcting, self-monitoring, self-organizing, self-replicating, or self-controlled functions. These capabilities are the result of the incorporation of technologies for sensing, actuation, coordination, communication, control, etc.”* [48, p. 3].

4.3 Cyber-Physical Systems

Hauser et al. [52] state that research on *Cyber-Physical Systems (CPS)* no longer takes place only in the disciplines of electronics and computer science, but also extends to other fields such as IS. Therefore, they describe a CPS as the extension of a legacy system with information technology [52]. Similarly abstract is the definition of Banerjee et al. [53], who describe CPS as *“systems that use the information from the physical environment, and in turn affect the physical environment”* [53, p. 283]. Furthermore, they list examples such as smart electricity grid and unmanned aerial vehicles [53]. Likewise, Gölzer et al. [6] argue that CPS are *“able to communicate with each other, to detect their environment, to interpret available data and to act on the physical world”* [6, p. 1]. They also emphasize the capabilities of self-control and self-optimization [6], while Gruettner et al. [54] describe CPS as *“intelligent networking of people, machines, and industrial processes, which in product components communicate with the production gear by embedded sensors”* [54, p. 1853].

Bradley and Atkins [55] state that CPS *“interface physics-based and digital world models”* [55, p. 60] and emphasize the benefits of integrating physical and computational models.

A formal definition is provided by Burmester et al. [56] describing a CPS as a *“finite state system consisting of several networked components, some of which may be cyber while others are physical”* [56, p. 3].

Akkaya et al. [57] identify the challenges of designing a Cyber-Physical System as *“complexity, heterogeneity, and multidisciplinary nature”* [57, p. 997], but avoid using a distinct definition. In addition, there are some articles that use the term CPS, but neither describe nor define it [58–62]. Other authors give examples such as smart grids [63, 64], Machine-to-Machine communication [65] and data centers [66], but also avoid clear definitions. However, most authors describe CPS basically as a conjunction of computation and physical processes, where there is a mutual influence through observation and control [67–73].

Böhmann et al. [36] build the bridge to Service Systems and explain that the availability of data and automation capabilities provided by Cyber-Physical Systems contribute to Service System innovation. Matzner and Scholta [47] also combine the CPS and Service Systems concepts and define: *"[CPS] are service systems that connect physical and cyber elements through global networks"* [47, p. 0].

Furthermore, Gunes et al. [5] summarize some aspects of different definitions and define CPS as *"complex, multi-disciplinary, physically-aware next generation engineered systems that integrate embedded computing technology (cyberpart) into the physical phenomena"* [5, p. 4244], where integration is achieved by the capabilities of *"observation, communication, and control [...] of the physical system"* [5, p. 4244].

Sanislav and Miclea [74] also recognize the variety of different definitions provided in the existing literature and list several, however, without synthesizing or providing their own.

Ribeiro et al. [75] and Wu et al. [73] emphasize the intelligence of such systems and characterize CPS as *"intelligent systems that are composed of digital virtual/cyber technologies, software, and physical components, and intelligently interact with other systems across information and physical interfaces"* [75, p. 6131]. Sampigethaya and Poovendran [76] consider CPS based on applications in aviation and describe mainly benefits and challenges. Also Sztipanovits et al. [77] and Yao et al. [78] focus mainly on challenges related to the integration of the various computational and physical elements of CPS.

Furthermore, Wan et al. [79] describe some characteristics of CPS such as *"cyber capability in every physical component"* [79, p. 1108], *close integration*, *"dynamically reorganizing/reconfiguring"* [79, p. 1108], and *"high degrees of automation"* [79, p. 1108].

We recommend following the definition of the majority of the authors and, thus, we provide an abstract definition: *"A Cyber-Physical System is an intelligent system connecting the physical and the digital/cyber world through influence and control using sensors and actuators"*.

4.4 Summary

This literature review shows that the concepts Service System, Smart Service System and Cyber-Physical System are not uniformly defined and also that the differentiation is not always clear. While most authors agree on Service Systems, Smart Service Systems and CPS in particular are not clearly defined.

By applying an open coding approach, properties of the examined concepts described in the articles are codified. Codes with similar characteristics are clustered and, thus, grouped together in categories [10]. Overall, we identify five categories of properties the concepts Service System, Smart Service System and Cyber-Physical System have in common. Table 1 depicts five identified categories components, attributes, actions, structure and boundaries. The categories components, attributes, and actions include a set of codes resulting from the different views of the articles being analyzed. We consider the most frequently occurring representatives for these three categories.

Table 1. Conceptualization of (Smart) Service Systems and Cyber-Physical Systems

	Service System	*Smart Service System*	*Cyber-Physical System*
key components	information, people, technology	data, people, technology	cyber part, sensors, actuators
key attributes	interaction, dynamic, adaptive	interaction, adaptive, learning, decision-making	interaction, intelligent, distributed
key actions	value creation	sensing, control	sensing, control
structure	complex, people-centered	complex, self-centered	complex, data-centered
boundaries	open, dynamic	open, dynamic	open, partially dynamic

The key components of all three concepts are frequently mentioned in the definitions within the articles and are also conceptually very clear, especially in the concepts of Service System and Smart Service System. For example, Service Systems and Smart Service systems both include *people* and *technology*, while in terms of Service Systems, the term *information* is very present, data is often referred to in Smart Service Systems. A CPS consists of a *cyber part* that provides computational capabilities, *sensors* collecting data, as well as *actuators*.

A variety of attributes are mentioned across all analyzed articles, however, only the key attributes are listed in Table 1. All three concepts emphasize the *interaction* between components, but also the interaction with the environment. Likewise, the attribute *adaptability* appears for all three concepts, although it is not mentioned as often in CPS definitions as the attribute *distributed*. In addition, the code *dynamic* is very common in Service Systems, while a CPS is particularly described as *intelligent* and Smart Service Systems is capable to *learn* and *make decisions*.

However, a small number of key actions are named, but the ones named are mentioned very frequently. Nearly every article defining a Service System names the goal of *creating value*. For Smart Service Systems and CPS, the actions are not quite as clean, but for both the two most common are *sensing* and *control*.

The structure of all three analyzed systems is described as a complex. In addition, Service Systems focus on people—both as component and user—while Smart Service Systems focus on the system itself and its purpose. CPS are often outlined as data-centered.

All three concepts are considered to be *open* systems. Furthermore, Service Systems and Smart Service Systems are able to change *dynamically*, while for CPS at least the physical part is fixed, but the components of the cyber part can also change dynamically.

5 Discussion

The analysis of the literature on the three concepts shows that Service Systems can be understood as socio-technical systems [29, 36–38, 48, 80]. In addition, a Smart Service Systems is a special kind of a Service System [7, 33, 50, 81]. CPSs, on the other hand, are referred to as a kind of Service System [47], but more often characterized as technical systems [5, 31, 82–84], which can thus be part of a socio-technical and, thus, part of a (Smart) Service System.

The analysis also shows that the need for information in Service Systems is enormous as it acts as a key component. The same applies to data in Smart Service Systems. This data, which can be further processed into information, can be collected by CPS. Thus, by enriching CPS with connectivity capabilities, the need for information / data of (Smart) Service Systems can be met. In addition, intelligent CPS can also serve as a social component to mimic the role of people.

Thus, the concepts Service System, Smart Service System and CPS are closely interlinked and, therefore, have similar characteristics. All concepts emphasize the interaction between humans and technology and the ability for multi-criteria decision-making. This leads to extremely complex and heterogeneous structures that can dynamically adapt over time.

In addition to components such as humans, technology or CPS, however, Service Systems themselves can also be components of Service Systems. This system-of-system property affects all three concepts. Thus, the system boundaries can be extended by parts of the environment, so that other systems arise.

Figure 2 depicts the interrelations of the three considered concepts as well as their connections to socio-technical system and system-of-systems concepts.

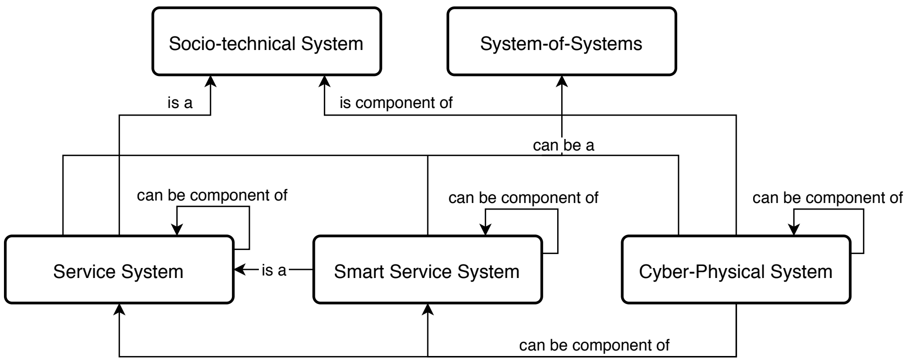


Figure 2. Interrelations of (Smart) Service Systems and Cyber-Physical Systems

6 Conclusion

The concepts of (Smart) Service Systems and Cyber-Physical Systems has been a re-occurring term in research and industry. Aiming for precise definitions, distinctions and similarities, we apply a thorough literature research and review 110 relevant articles. As a result, we show that especially the concepts Smart Service System and Cyber-Physical System are often used in a similar context in different disciplines. The concepts include similar facets and characteristics. However, our research reveals some cases of inconsistent definitions, especially for the concepts of Smart Service Systems and Cyber-Physical Systems. For clarification, we derive suitable definitions from literature and fuse them in a conceptualization. These definitions and concepts may assist researchers in the understanding of the terms and their relationships.

Our work is limited to literature originating mainly from the fields of Information Systems, Service Science and Computer Science community. Furthermore, it can remain subjective as to whether a definition is more suitable than another to understand broader concepts. To address this, we based our research on occurrences in related articles, but cannot account for all articles across all disciplines. Moreover, the identified characteristics are not validated concerning their completeness and meaning within different disciplines. In total, this work sets a foundation for researchers and practitioners to understand the concepts consistently and, accordingly, to push for new research and development in these areas with the same terminology in mind to avoid misunderstandings.

References

1. Chen, H., Chiang, R.H.L., Storey, V.C.: Business Intelligence And Analytics: From Big Data To Big Impact. Mis Q. 36, 1165–1188 (2012).
2. Davenport, T., Harris, J.: Competing on Analytics: Updated, with a New Introduction: The New Science of Winning. Harvard Bus. Press. (2017).
3. Spohrer, J., Maglio, P.P., Bailey, J., Gruhl, D.: Steps toward a science of service systems. Computer (Long. Beach. Calif). 40, 71–77 (2007).
4. Maglio, P.P.: Editorial Column — Smart Service Systems. Serv. Sci. 6, i–ii (2014).
5. Gunes, V., Peter, S., Givargis, T., Vahid, F.: A survey on concepts, applications, and challenges in cyber-physical systems. KSII Trans. Internet Inf. Syst. 8, 4242–4268 (2014).
6. Gölzer, P., Cato, P., Amberg, M.: Data Processing Requirements of Industry 4.0- Use Cases for Big Data Applications. In: Proceedings of the European Conference On Information Systems. pp. 1–13 (2015).
7. Barile, S., Polese, F.: Smart Service Systems and Viable Service Systems: Applying Systems Theory to Service Science. Serv. Sci. 2, 21–40 (2010).
8. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. Proc. 17th Eur. Conf. Inf. Syst. Verona. 2206–2217 (2009).
9. Cooper, H.M.: Organizing knowledge synthesis: a taxonomy of literature reviews. Knowl. Soc. 1, 104–126 (1988).
10. Saldaña, J.: The Coding Manual for Qualitative Researchers. Cambridge University Press, Cambridge (2009).
11. Standards Coordinating Committee of the Computer Society of the IEEE: IEEE Standard Glossary of Software Engineering Terminology, (1990).
12. Boulding, K.E.: General systems theory: The skeleton of science. Manage. Sci. 2, 197–208 (1956).
13. Sagawa, T.: Thermodynamics of Information Processing in Small Systems. Springer Japan, Tokyo (2013).
14. Bertalanffy, L. von: An Outline of General System Theory. Br. J. Philisophy Sci. 1, 134–165 (1950).
15. Baxter, G., Sommerville, I.: Socio-technical systems: From design methods to systems engineering. Interact. Comput. 23, 4–17 (2011).

16. Bostrom, R.P., Heinen, J.S., Heinen, J.S.: MIS Problems and Failures : A Socio- Technical Perspective Part I : The Causes. MIS Q. 1, 17–32 (1977).
17. Cartelli, A.: Socio-Technical Theory and Knowledge Construction: Towards New Pedagogical Paradigms? Issues Informing Sci. Inf. Technol. 4, 1–14 (2007).
18. Gideon, J.M., Dagli, C.H., Miller, A.: Taxonomy of Systems-of-Systems. In: Conference on Systems Engineering Research, Institute of Electrical and Electronics Engineers (IEEE) (2005).
19. Maier, M.W.: Architecting Principles for Systems of Systems. Syst. Eng. 1, 267–284 (1998).
20. Maglio, P.P., Spohrer, J.: Fundamentals of service science. J. Acad. Mark. Sci. 36, 18–20 (2008).
21. Maglio, P.P., Vargo, S.L., Caswell, N., Spohrer, J.: The service system is the basic abstraction of service science. Inf. Syst. E-bus. Manag. 7, 395–406 (2009).
22. Baekgaard, L.: Service Scenarios - A Socio-Technical Approach to Business Service Modeling. In: Proceedings of the European Conference on Information Systems (2009).
23. Edvardsson, B., Tronvoll, B., Gruber, T.: Expanding understanding of service exchange and value co-creation: a social construction approach. J. Acad. Mark. Sci. 39, 327–339 (2011).
24. Jaakkola, E., Alexander, M.: The Role of Customer Engagement Behavior in Value Co-Creation : A Service System Perspective. J. Serv. Res. 17, 247–261 (2014).
25. Zhou, Z., Lin, Y., Yue, F.: Service-Dominant Logic for Exploring Modular Business Service System. In: Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics. pp. 108–112 (2014).
26. Kleinschmidt, S., Burkhard, B., Hess, M., Peters, C., Leimeister, J.M.: Towards design principles for aligning human-centered service systems and corresponding business models. In: Proceedings of the International Conference on Information Systems. pp. 2–11 (2016).
27. Kleinschmidt, S., Peters, C., Leimeister, J.M.: ICT-Enabled Service Innovation in Human-Centered Service Systems: A Systematic Literature Review. In: Proceedings of the International Conference on Information Systems. pp. 1–18 (2016).
28. Ralyté, J., Khadraoui, A., Léonard, M.: Designing the Shift from Information Systems to Information Services Systems. Bus. Inf. Syst. Eng. 57, 37–49 (2015).
29. Eaton, B., Elaluf-calderwood, S., Sørensen, C., Eaton, B.: Distributed tuning of boundary resources: the case of Apple's iOS service system. MIS Q. 39, 217–243 (2015).
30. Knote, R., Blohm, I.: It's not about having Ideas - It's about making Ideas happen! Fostering Exploratory Innovation with the Intrapreneur Accelerator. In: Proceedings of the European Conference on Information Systems (2016).
31. Herterich, M.M., Eck, A., Uebernickel, F.: Exploring How Digitized Products Enable Industrial Service Innovation – An Affordance Perspective. In: Proceedings of the European Conference on Information Systems (2016).
32. Brust, L., Antons, D., Breidbach, C.F., Salge, T.O.: Service-Dominant Logic and Information Systems Research: A Review and Analysis Using Topic Modeling. In: Proceedings of the International Conference on Information Systems (2017).
33. Spohrer, J., Siddike, M.A.K., Kohda, Y.: Rebuilding Evolution: A Service Science Perspective. In: Proceedings of the Hawaii International Conference on System Sciences. pp. 1663–1672 (2017).
34. Kleinschmidt, S., Peters, C.: Towards an integrated evaluation of human-centered service systems and corresponding business models: A systems theory perspective. In: Proceedings of the ECIS. pp. 3060–3070 (2017).
35. Lintula, J., Tuunanen, T., Salo, M.: Conceptualizing the Value Co-Destruction Process for Service Systems : Literature Review and Synthesis. In: Proceedings of the Hawaii International Conference on System Sciences. pp. 1632–1641 (2017).

36. Böhmann, T., Leimeister, J.M., Möslein, K.: Service Systems Engineering: A Field for Future Information Systems Research. Bus. Inf. Syst. Eng. 6, 73–79 (2014).
37. Dörbecker, R., Tilo Böhmann: FAMouS – Framework for Architecting Modular Services. In: Proceedings of the International Conference on Information Systems. pp. 1–18 (2015).
38. Li, M.M., Peters, C.: Mastering Shakedown Through The User: The Need for User-Generated Services In Techno Change. In: Proceedings of the European Conference on Information Systems (2016).
39. Höckmayr, B., Roth, A.: Design of a Method for Service Systems Engineering in the Digital Age. In: Proceedings of the International Conference on Information Systems. pp. 1–23 (2017).
40. Motta, G., You, L., Sacco, D., Miceli, G.: Mobility Service Systems: guidelines for a possible paradigm and a case study. In: Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics. pp. 48–53 (2014).
41. Alter, S.: Service System Fundamentals: Work System, Value Chain, and Life Cycle. IBM Syst. J. 47, 1–25 (2008).
42. Alter, S.: Metamodel for Service Design and Service Innovation: Integrating Service Activities, Service Systems, and Value Constellations. In: Galletta, D.F. and Liang, T.-P. (eds.) Proceedings of the International Conference on Information Systems. Association for Information Systems (2011).
43. Alter, S.: Service System Axioms that Accept Positive and Negative Outcomes and Impacts of Service Systems. In: Proceedings of the International Conference on Information Systems. pp. 1–21 (2017).
44. Alter, S.: Answering Key Questions for Service Science. In: Proceedings of the European Conference on Information Systems. pp. 1822–1836 (2017).
45. Blohm, I., Haas, P., Peters, C., Jakob, T., Leimeister, J.M.: Managing Disruptive Innovation through Service Systems – The Case of Crowdlending in the Banking Industry. In: Proceedings of the International Conference on Information Systems. pp. 1–15 (2016).
46. Dörbecker, R., Tokar, O., Böhmann, T.: Deriving Design Principles for Improving Service Modularization Methods - Lessons Learnt from the Complex Integrated Health Care Service System. In: Proceedings of the European Conference on Information Systems. pp. 0–15 (2015).
47. Matzner, M., Scholta, H.: Process Mining Approaches to Detect Organizational Propoerties in Cyber-Physical Systems. In: Proceedings of the European Conference On Information Systems (2014).
48. Medina-Borja, A.: Editorial Column—Smart Things as Service Providers: A Call for Convergence of Disciplines to Build a Research Agenda for the Service Systems of the Future. Serv. Sci. 7, ii–v (2015).
49. Beverungen, D., Müller, O., Matzner, M., Mendling, J., Vom Brocke, J.: Conceptualizing smart service systems. Electron. Mark. 1–12 (2017).
50. Maglio, P., Lim, C.-H.: Innovation and Big Data in Smart Service Systems. J. Innov. Manag. 4, 11–21 (2016).
51. Santo, M. De, Pietrosanto, A., Napoletano, P., Carrubbo, L.: Knowledge based service systems. Syst. Theory Serv. Sci. Integr. Three Perspect. a New Serv. Agenda. (2011).
52. Hauser, M., Günther, S.A., Flath, C.M., Thiesse, F.: Designing Pervasive Information Systems : A Fashion Retail Case Study. In: Proceedings of the International Conference on Information Systems. pp. 1–16 (2017).
53. Banerjee, A., Venkatasubramanian, K.K., Mukherjee, T., Gupta, S.K.S.: Ensuring safety, security, and sustainability of mission-critical cyber-physical systems. In: Proceedings of the IEEE. pp. 283–299 (2012).

54. Gruettner, A., Richter, J., Basten, D.: Explaining the Role of Service-Oriented Architecture for Cyber-Physical Systems By Establishing Logical Links. In: Proceedings of the ECIS. pp. 1853–1868 (2017).
55. Bradley, B.J.M., Atkins, E.M.: Toward Continuous State–Space Regulation of Coupled Cyber–Physical Systems. In: Proceedings of the IEEE. pp. 60–74 (2012).
56. Burmester, M., Magkos, E., Chrissikopoulos, V.: Modeling Security in Cyber-Physical Systems. Int. J. Crit. Infrastruct. Prot. 5, 118–126 (2012).
57. Akkaya, I., Derler, P., Emoto, S., Lee, E.A.: Systems Engineering for Industrial Cyber – Physical Systems Using Aspects. In: Proceedings of the IEEE. pp. 997–1012 (2016).
58. Janiesch, C., Diebold, J.: Conceptual Modeling of Event Processing Networks. In: Proceedings of the European Conference On Information Systems (2016).
59. Jaskolka, J., Villasenor, J.: Identifying Implicit Component Interactions in Distributed Cyber-Physical Systems. In: Proceedings of the HICSS. p. 10 (2017).
60. Jin, J., Gubbi, J., Marusic, S., Palaniswami, M.: An Information Framework for Creating a Smart City Through Internet of Things. IEEE Internet Things J. 1, 112–121 (2014).
61. Tabuada, P., Caliskan, S.Y., Rungger, M., Majumdar, R.: Towards robustness for cyber-physical systems. IEEE Trans. Automat. Contr. 59, 3151–3163 (2014).
62. Venkitasubramaniam, P., Yao, J., Pradhan, P.: Information-Theoretic Security in Stochastic Control Systems. In: Proceedings of the IEEE. pp. 1914–1931 (2015).
63. Siaterlis, C., Genge, B.: Theory of evidence-based automated decision making in cyber-physical systems. In: 2011 IEEE International Conference on Smart Measurements of Future Grids (SMFG) Proceedings. pp. 107–112 (2011).
64. Yu, X., Xue, Y.: Beyond Smart Grid—A Cyber–Physical–Social System in Energy Future. In: Proceedings of the IEEE. pp. 2290–2292 (2017).
65. Gharbi, G., Guermouche, N., Monteil, T.: Temporal Verification of Mobile Publish / Subscribe Machine-to-Machine Communications. In: Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks. pp. 2–4 (2014).
66. Parolini, L., Sinopoli, B., Krogh, B.H., Wang, Z.K.: A cyber-physical systems approach to data center modeling and control for energy efficiency. In: Proceedings of the IEEE. pp. 254–268 (2012).
67. Derler, P., Lee, E.A., Sangiovanni Vincentelli, A.: Modeling cyber-physical systems. In: Proceedings of the IEEE (2012).
68. Han, S., Xie, M., Chen, H.H., Ling, Y.: Intrusion detection in cyber-physical systems: Techniques and challenges. IEEE Syst. J. 8, 1049–1059 (2014).
69. Lee, E.A.: Cyber Physical Systems: Design Challenges. (2008).
70. Nuzzo, P., Sangiovanni-Vincentelli, A.L., Bresolin, D., Geretti, L., Villa, T.: A Platform-Based Design Methodology With Contracts and Related Tools for the Design of Cyber-Physical Systems. In: Proceedings of the IEEE. pp. 2104–2132 (2015).
71. Poovendran, R.: Cyber-physical systems: Close encounters between two parallel worlds. In: Proceedings of the IEEE. pp. 1363–1366 (2010).
72. Rajhans, A., Bhave, A., Ruchkin, I., Krogh, B.H., Garlan, D., Platzer, A., Schmerl, B.: Supporting Heterogeneity in Cyber-Physical Systems Architectures. In: IEEE Transactions on Automatic Control. pp. 3178–3193 (2014).
73. Wu, F.J., Kao, Y.F., Tseng, Y.C.: From wireless sensor networks towards cyber physical systems. Pervasive Mob. Comput. 7, 397–413 (2011).
74. Sanislav, T., Miclea, L.: An agent-oriented approach for cyber-physical system with dependability features. In: Proceedings of 2012 IEEE International Conference on Automation, Quality and Testing, Robotics. pp. 356–361 (2012).

75. Ribeiro, F., Rettberg, A., Pereira, C.E., Soares, M.S.: A Model-Based Engineering Methodology for Requirements and Formal Design of Embedded and Real-Time Systems. In: Proceedings of the HICSS. pp. 6131–6140 (2017).
76. Sampigethaya, K., Poovendran, R.: Aviation cyber-physical systems: Foundations for future aircraft and air transport. In: Proceedings of the IEEE. pp. 1834–1855 (2013).
77. Sztipanovits, J., Koutsoukos, X., Karsai, G., Kottenstette, N., Antsaklis, P., Gupta, V., Goodwine, B., Baras, J., Wang, S.: Toward a science of cyber-physical system integration. In: Proceedings of the IEEE. pp. 29–44 (2012).
78. Yao, J., Xu, X., Liu, X.: MixCPS : Mixed Time / Event-Triggered Architecture of Cyber – Physical Systems. In: Proceeding of IEEE (2016).
79. Wan, J., Chen, M., Xia, F., Li, D., Zhou, K.: From machine-to-machine communications towards cyber-physical systems. Comput. Sci. Inf. Syst. 10, 1105–1128 (2013).
80. Qiu, R.G.: Computational Thinking of Service Systems: Dynamics and Adaptiveness Modeling. Serv. Sci. 1, 42–55 (2009).
81. Lim, C., Maglio, P.P.: Data-Driven Understanding of Smart Service Systems Through Text Mining. Serv. Sci. 10, 154–180 (2018).
82. Huang, X., Dong, J.: Reliable Control Policy of Cyber-Physical Systems Against a Class of Frequency-Constrained Sensor and Actuator Attacks. IEEE Trans. Cybern. 1–8 (2018).
83. Jirkovsky, V., Obitko, M., Marik, V.: Understanding Data Heterogeneity in the Context of Cyber-Physical Systems Integration. IEEE Trans. Ind. Informatics. 13, 660–667 (2017).
84. Kang, W., Kapitanova, K., Son, S.: RDDS: A real-time data distribution service for cyber-physical systems. IEEE Trans. Ind. Informatics. 8, 393–405 (2012).

Developing an Industrial IoT Platform – Trade-off between Horizontal and Vertical Approaches

Louisa Schermuly, Maximilian Schreieck, Manuel Wiesche, and Helmut Krcmar

Technical University of Munich, Chair for Information Systems, Munich, Germany
louisa.schermuly@tum.de,{maximilian.schreieck,
wiesche,krcmar}@in.tum.de

Abstract. Demands for a digitalized, connected, and smart production provide a fertile ground for industrial Internet-of-Things (IIoT) platforms to arise within the manufacturing industry (e.g., Siemens MindSphere, AXOOM Smart Enterprise, FORCAM FORCE). Nevertheless, many companies struggle to successfully kick-off platform ecosystems. Information Systems (IS) literature is of limited help, because insights on managing platform ecosystems are mostly derived from successful examples in the business-to-consumer (B2C) context. To better understand the challenging situation of companies in the emerging IIoT environment, we conducted an in-depth case study at a prospective platform provider. Insights gained through interviews and engagement in the field uncovered a tension between a horizontal platform strategy and vertical integrated solutions as a central challenge for companies aiming to launch an IIoT-platform in the market. By conceptualizing this trade-off, its causes along with related benefits and challenges, we add to existing literature on platform governance and launch strategies.

Keywords: Software Platforms, Industrial Internet-of-Things, CPS Platforms, Platform Development, Exploratory Case Study

1 Introduction

In the course of the ongoing fourth industrial revolution (‘Industry 4.0’), advanced and innovative information and communication technologies (ICT) are being introduced to logistic and manufacturing companies to facilitate more automated and intelligent processes. Within this context, the importance of service-based, real-time enabled cyber physical systems (CPS) was not only identified in theory [1]. In practice, this constitutes in the emergence of a vast number of Industrial Internet of Things (IIoT) platforms, which enable the connection of machines, devices, materials and sensors to lay the foundation for a smart production system [2, 3].

Information systems (IS) researchers have studied the concept of platforms for several years. A common definition is given by Tiwana, who defines a digital platform as “a software-based product or service that serves as a foundation on which outside parties can build complementary products or services” [4, p. 5]. A growing body of research in this field investigated the concept of platforms from different perspectives.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Some studies follow an economic perspective investigating two-sided market characteristics of platforms like network externalities, multi-homing behavior or pricing strategies [5, 6, 7], others focus on technological architectures, open decisions and a platform's innovation potential by efficient value co-creation [8, 9, 10, 11].

Even though we find extensive literature about digital platforms, the applicability in the emerging market of IIoT platforms is limited, as most theories are based on findings from platform ecosystems operating in a business-to-consumer (B2C) context (e.g., web browsers, personal computer and mobile operating systems, video consoles, Facebook, Google search engine etc.) [12, 13, 14, 15]. Business-to-business (B2B) platforms – or B2B products and services in general – differ not only in their technological complexity. Also, different market characteristics such as size, fragmentation, and business relationships entail divergent requirements for platform strategies [e.g., 16]. As market structure shapes a platform's emergence process, the emergence processes observable in the context of Industry 4.0 differ from what has been researched in existing literature. The concept of platform development has also often been studied from a broad industry perspective [17, 18], or with focus on mature platforms [19]. Concerning the start-up phase of a platform, IS researchers agree that most successful platforms emerged from existing products in an evolutionary process [4, 20]. However, literature about this emergence process and involved actions and decisions of the platform owner remains limited.

In practice, we note that many prospective platform owners find themselves in a situation where market demands a vertical (product-oriented) and a horizontal (platform-oriented) strategy simultaneously. Literature has rarely addressed the existence and related challenges of such a hybrid strategy for platform emergence. Instead, literature highlights the difference between products or services and platforms in a sense that strategic decisions being made can have contrasting implications. Thus, managers need to decide early, whether to follow a vertical product or a horizontal platform strategy [4, 21, 22].

The purpose of this study is to gain a deeper understanding of a platform's emergence process from the perspective of a platform owner. More particularly, we focus on the platform owner's trade-off between following a vertical vs. a horizontal approach. To understand why this situation arises and how platform owners deal with this trade-off, we defined the following research questions: *Which factors cause the trade-off between following a vertical vs. a horizontal approach? How do platform owners try to reconcile these contrary requirements and what are benefits or challenges involved?*

In the remainder of this paper, we will first provide a brief overview of this paper's theoretical background. Then, we present the design of our exploratory in-depth case study [23, 24]. In order to answer the research questions defined above, we followed the ideas of engaged scholarships [25] and conducted semi-structured interviews at a medium-sized company currently developing an IIoT platform. Based on our findings, we developed a framework explaining the trade-off's circumstances, resulting strategy chosen by the case company as well as its consequences. Further, we discuss our findings against the background of the platform's B2B context. Finally, we present limitations of the study as well as implications for future work.

2 Theoretical Background

The platform concept has been researched in several fields of studies resulting in different definitions and views on the platform phenomenon [8]. Within IS two strands of literature can be identified [26]. The *market-oriented perspective* emphasizes a platform's characteristic to act as an intermediary between two or multiple sides, whereas researchers following the *technology-oriented perspective* study a platform's architecture and its capabilities to facilitate value co-creation and foster innovation. In theory, only a few researches consider both perspectives within their studies [26], which diminishes applicability of existing theories.

Within this paper, the term platform is used to describe a construct encompassing both platform perspectives. These encompassing platforms correspond to the notion of 'industry platforms' found in the typology of Gawer [20]. Furthermore, it matches the widespread definition by Tiwana, who describes a platform as 'a software-based product or service that serves as a foundation on which outside parties can build complementary products or services' [4, p. 5]. Common examples are the mobile operating systems of Google and Apple. Both platforms provide open application interfaces (APIs) allowing complementors (often referred to as third-party developers) to reuse core functionalities and innovate upon these platforms by developing complementary products and services in form of apps. In addition, Apple and Google offer a marketplace where app developers can sell their apps to end users. Thus, the platforms act as an intermediary between two sides.

2.1 Shifting Towards Horizontally Organized Platform Ecosystems

Literature has acknowledged that more and more markets shift towards a platform-centric market structure. In many cases this shift can be traced back to the digitalization of the industry involved [27, 28]. In addition to the increased software embedding, ubiquity of wireless data networks, IoT and packetization, and the need for deepening specialization are drivers for this transformation [4]. Within affected markets, vertical integrated product offerings get replaced by a layered, modular architecture of digital technologies providing the ground for platforms to rise [18, 27]. In this context, the development and provision of an integrated solution (e.g. from sensor to specific end user application) by a single or small group of firms is referred to as the vertical approach in this paper. In contrast, a company following a horizontal approach decided to focus on a single layer and thus the development of a sophisticated and comprehensive open IIoT platform. From the market perspective, the occurrence of these platforms results in a greater specialization of single firms developing complementary solutions and unlocks an increased pace of innovation in the industry [4, 10].

In this setting, scholars differentiate between *inter*platform and *intra*platform competition. The overall market competition shifted towards a competition between platform ecosystems [28, 29], in which incumbent platforms strengthen their position automatically through network effects [30]. Additionally, high innovation capabilities of an ecosystem make it hard for closed vertical products to compete. In contrary, intraplatform competition refers to the competition between complementors [31]. This competition ensures quality and price regulation [20], even in the common scenario of a market with one leading platform ('Winner-takes-all') [32].

This ecosystem competition requires the management of firms to adopt a new mindset [4]. Stepping into the role of a platform owner means that value creation does no longer happen internally or in few strategic partnerships [33]. The new focus of platform owners is to orchestrate a big, loosely coupled ecosystem of firms for efficient value co-creation [34, 35, 36, 37]. Innovation is not planned anymore, but rather emerges in the ecosystem [4]. Nevertheless, through the provision of appropriate boundary resources [14, 15, 38, 39] or different pricing strategies [6] a platform owner can still foster or hamper the emergence of innovation. Further platform-related management challenges are summarized in the 'Four Levers of Platform Leadership' by Cusumano and Gawer [40]. They include decisions about scope, product technology, relationships with external complementors and internal organization. Referring to these differences, literature highlights the importance of an early and clear management decision, whether to follow a platform strategy or to offer a vertical integrated (product) solution [4, 21, 22]. Otherwise, it 'can result in dangerous strategic confusion' [21, p. 69], as decisions being made in platform competition differ from those, a firm makes while pursuing a product strategy. Nevertheless, other scholars consider a hybrid business model as useful in the transition phase between product and platform business [41], but also acknowledge that especially small firms may not be able to afford such a hybrid strategy [27].

2.2 Platform Emergence from the Perspective of a Platform Owner

For prospective platform owners, it is necessary to detect the shift towards a layered market, determine their role in this ecosystem (platform leader or complementor) and adjust the strategy accordingly [18]. Most successful platforms arise out of a product or service by including a second market side [4]. Thereby, platform potential of a product can be identified by assessing the existence of a second group interested in a cost-effective interaction with existing customers, the potential for cross-side network effects and/or the existence of unexploited long-tails in the prospective market [4]. Others suggest that the following two conditions should be fulfilled: (1) the product, service or technology needs to provide essential functionalities solving an existing problem of an industry and (2) it offers easy-to-use interfaces to develop complementary services [21]. However, adding a second side can also be interpreted as a process of opening the platform and thereby giving up control. This is also reflected in the evolutionary perspective on platforms by Gawer [20], where internal product platforms evolve into industry platforms.

In sum, the theories and strategies provided in literature are not sufficient to understand the specific steps and actions of platform owners during the emergence of platform development. Furthermore, the mostly linear emergence models proposed in literature do not consider, that market demands may require a company to pursue a vertically integrated product strategy and a horizontally platform strategy simultaneously. However, especially in the emerging market for IIoT solutions, the existence of these contrary requirements is a common situation for example for providers of IoT technologies when extending their solutions by complementary software offers.

3 Research Design

To explore the process of platform emergence from the perspective of a platform owner, we conducted a single case study [23, 24]. Considering the complexity of a platform's emergence process, it is reasonable to study the phenomenon in its real-life context. Therefore, we did not only conduct semi-structured interviews, but also gained valuable insights through engagement in the field [25]. Furthermore, the 'how' and 'why' research questions confirm the suitability of a case study research approach, which was conducted in an exploratory manner [24]. In this sense, the case study is also designed to reveal yet unknown events and challenges in the process of platform development to later build theory upon [23]. Overall, the case study was carried out in an iterative process of data collection and analysis [23, 24, 42].

The case company being studied (referred to as *HardwareProvider* in the following) is a medium-sized, young company based in Germany. As many other players in the field, the company's background lays in the hardware business [43], so that a software platform was primarily developed as a value-added service for their customers from the manufacturing and logistic industry. The platform should be designed open to raise an ecosystem around it and thereby foster innovation. Thus, the case company is currently facing the tension between following a vertically integrated product strategy and a horizontally open platform strategy. In this regard, *HardwareProvider* is an interesting case for two reasons. First, being in an early stage of its platform development allows observing strategical decisions while being made. Second, the limited business maturity and the company's small size simplify the otherwise very complex environment of an IIoT platform.

For **data collection**, we relied on primary and secondary data. We conducted seven semi-structured expert interviews with employees having a key role in the development of the company's platform offer [44]. Table 1 provides an overview of the interviewee's job responsibilities, date of interviews and the interview length. The respective interview partners (IP) got informed about the interview and its purpose beforehand. In the manner of an iterative study design, questions in the first interviews were rather broad to get an overview of the platform emergence and interview partners for further interviews [42]. In general, topics covered during the interviews included not only the technical development, but also strategical decision making related to platform design, marketing and related partnerships. Due to the fact, that each interview partner had a different role in the process of platform development, not every interviewee was asked the same questions. Besides interviews, additional primary and secondary data was accessible due to our engagement in the field over a period of five months. As our single case study at *HardwareProvider* follows the ideas of engaged scholarship [25], informal conversations and collaboration with key stakeholders from the case company took already place in an early stage to build a practical pre-understanding, formulate relevant research questions and design the study accordingly. In a later stage, additional data sources like marketing material, internal product roadmaps and project documentation, were used to validate findings from the interviews. Furthermore, the attendance of internal and external meetings and workshops allowed us to include perspectives from additional stakeholders not being interviewed directly (e.g. partners and customers).

Table 1. Summary of interviews (ID, date, interviewee, length)

ID	*Date*	*Interviewee position*	*Length*
IP1	19.06.18	Founder and managing director of *HardwareProvider*	50 min
IP2	21.06.18	Managing director of *HardwareProvider*	36 min
IP3	10.07.18	Head of project engineering	48 min
IP4	17.07.18	Head of development	55 min
IP5	18.07.18	Product and project manager	41 min
IP6	20.07.18	Sales and partner manager	49 min
IP7	25.07.18	Founder and managing director of *HardwareProvider* (follow-up)	39 min

For **data analysis**, we applied procedures of the grounded theory methodology [42, 45, 46]. We used the bottom-up coding technique to analyze qualitative data and find theoretical concepts. In a first step of open coding, we associated about 340 concepts to more than 450 interview quotes. Then, we grouped codes by defining emerging (sub-)categories (Table 2). During axial coding, we adapted the *Six C* coding family to relate identified categories to each other [47]. Furthermore, we wrote theoretical memos to record ideas raised during coding [42]. As a last step, we applied selective coding to establish the core category and identify all relevant categories to describe the phenomenon in more detail and to build a theoretical model.

Table 2. Illustration of coding scheme

Interview statement[1]	**Concepts**	**Categories**
"The past has shown that we did not do completely wrong, when following a project-driven development strategy. Because we knew, that the market needed the feature or this functionality[1,2]*. We could develop the feature in cooperation with the customer from industry*[2]*. Thus, also tap into a lot know-how*[2,3]*. But we also tried to develop new functionalities independently, which was also driven out of a new product management department*[4]*." [IP5]*	1. Minimizing risk through customer-oriented feature development 2. Cooperative feature development with customer 3. Building up know-how through cooperative development 4. New product management department	▪ Sources of platform features ▪ Project-driven development ▪ Independent feature development

[1] All quotes have been translated from German

4 Results and Interpretation

Within our analysis we did not only gain a deeper understanding of *HardwareProvider's* journey of becoming a platform provider, but we also identified the trade-off between horizontal and vertical platform development as a major issue faced by a platform owner during platform emergence. Therefore, we further studied the issue by investigating context, causes, (pre-)conditions, undertaken strategy and consequences. The established model is depicted in Figure 1 and will be discussed in the following.

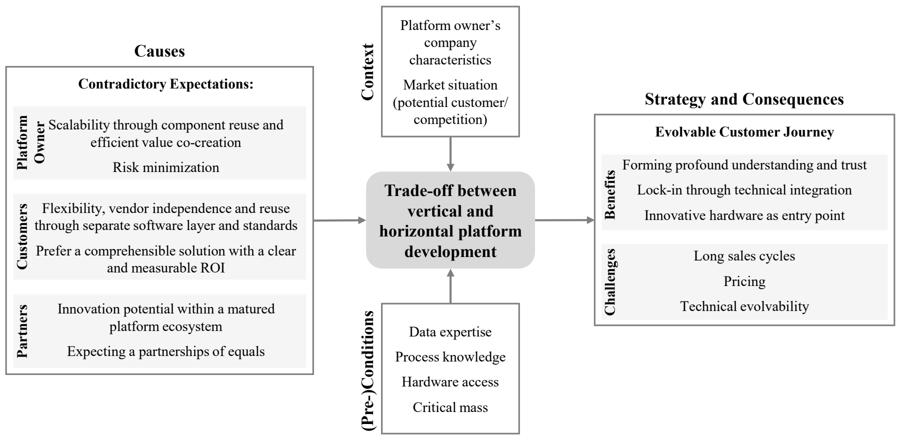


Figure 1. The challenging platform development environment

4.1 Emergence of the Horizontal vs. Vertical Trade-off

The case study helped to understand the circumstances, which brought the prospective platform owner into the trade-off between vertical product and horizontal platform development. The overall situation appears rather complex, though we defined three clusters shaping this trade-off.

Causes: The emergence of the trade-off is driven by anticipated consequences of internal and external stakeholders. On the one hand, the *platform owner*, especially members of the development and project management team, would benefit from focusing on a generic horizontal approach. Such an approach allows for more scalable project business through component reuse, as well as the possibility for efficient value co-creation by the customers themselves or external partners. On the other hand, customer-oriented implementations of vertical end-to-end use cases minimize the risk of implementing features not required by the market and simplifies marketing and sales.

From the *customer perspective*, a comprehensive, company-wide open platform for its data may be desirable as it can be used to set standards. Thus, flexibility and vendor independence for the implementation of vertical use cases can be achieved. This would in turn strengthen the customer's position in future purchasing processes: *'But the [customer's] strategy was, to not be dependent on a specific manufacturer, but to establish a layered model, in which vendors of every layer [e.g. hardware, software,* *application] can be easily replaced.'* [IP3] However, although big players and a few progressive medium-sized companies already recognized the potential of a horizontally, layered architecture, many customers, especially employees from operational departments, prefer use-case oriented solutions with a clear and measurable Return-on-Investment (ROI). Especially in manufacturing companies these stakeholders have an important role, as an interview partner mentioned: *'There were mainly the production managers, who wanted that a specific use case is tailored and targeted well-functioning. And as with any manufacturing company, production [departments] have a prominent position and weight […].'* [IP3]

Lastly, *potential partners* have contradictory expectations as well. While a horizontally matured platform approach would provide them with a fertile ground for innovation, they still expect a close partnership of equals with shared product portfolio often requiring partner-specific vertical use-case implementations. Although these contradictory partner requirements are mainly attributed to the variety of partner types, even within single partner companies, expectations contradict to a certain degree.

Context: The trade-off must be seen in the context it emerged in. On the one hand, the popularity of trends like 'Industry 4.0' or concepts of the 'Smart Factory' not only raised interest for innovative solutions, but also had a positive impact on the customers' or potential partners' willingness to invest. On the other hand, companies in the production and logistics industry are still rather conservative. That is, management staff often belongs to elderly generations lacking a well-versed technological understanding, thus struggling to recognize technological potential. This requires educational work, which in turn is easier when selling vertical end-to-end solutions. Furthermore, the market is flooded by a high number of (I)IoT platforms. In this context, potential customers get confused and a critical attitude towards the platform term is clearly noticeable, as it was pictured by an interview partner: *'[…] there are platforms like sand by the sea. If you go to the customer and you use the word platform, then most say: Please leave immediately, I can't, I really can't hear this [word] anymore.' [IP2]* Hence, a clear value proposition in form of an implemented vertical end-to-end use case is even more important for those customers.

However, not only the target market caused the trade-off being faced, but also the current situation of the platform owner. Conducting interviews at *HardwareProvider* it became clear, that especially the company characteristics like age, size and background are closely intertwined with the trade-off. To begin with, limited resources in the development, sales, partner management as well as the product and project engineering department intensify the trade-off, because a simultaneous realization of both concepts is not feasible. As the company was only founded a few years ago, it had only a small potential to exploit existing (hardware) customer base for upselling. In addition, due to a limited brand recognition, customers expect close customer contact, commitment by the platform owner and a clear value proposition: *'Well, as a company with our name [little prominence] and without any other relationships, we don't have the opportunity to just bluntly sell a platform, which may not have such a huge added value yet, because there are only few applications on it.'* [IP1]

Also for *HardwareProvider,* such a close collaboration is desirable, as they can acquire relevant domain and process knowledge. Lastly, the company went through a transformation of the business focus as it was operating in the hardware sector before. The shift towards software business was challenging as the company lacked not only technical expertise in this field, but also experience concerning software sales processes and partnerships. By offering vertical end-to-end solutions combining software platform and own hardware products, *HardwareProvider* can leverage their existing resources to obtain a competitive advantage: *'[…] that we had a great deal of know-how in the field of [this type of] data and we were able to develop the right platform for it. […] That is, I need to differentiate my offer by knowing the data well, understand related use cases and then having a platform for it.'* [IP2]

(Pre-)Conditions: To pursue either of the two approaches a set of conditions need to be fulfilled. A vertical end-to-end solution requires access to hardware components as well as expertise about the created data (transportation, structure, content). In this context, one interview partner explained, why interface standards are only of limited use: '*[...] but if you are able to understand it - that is a different problem. The standardization of [data] transportation […] is relatively easy. And this is, what most concentrate on, when implementing standard interfaces. And the other [aspects] are often forgotten. Things like content structure. But even if the content is clear, then you still have things like units and coordinate systems. […] standard definitions leave questions, for example how certain information like battery status is communicated. That isn't standardized. And every supplier does it a bit different. Thus, it is not completely standardized anymore.'* [IP4] In addition, process knowledge is necessary to identify relevant use cases for implementation. While for the implementation of vertical use cases mostly specific data expertise is required, for successfully following the horizontal approach an even broader data expertise is necessary. Furthermore, a critical mass is relevant for making reasonable investments in the development of platform related features like an app store, user management, interface management, or automatic licenses.

4.2 Resulting Strategy and Consequences

Being confronted with this trade-off, *HardwareProvider* established a strategy to combine the two approaches by pursuing an **evolvable customer journey**. Thereby, the concept of a customer journey describes different stages of the actual purchasing cycle, as well as the relation to previous and future purchase experiences [48]. We denote the customer journey as evolvable, because of the customer's shift of interests: from requiring a vertical integrated solution towards requesting horizontal platform characteristics.

Starting with small, timely restricted projects as a proof-of-concept, *HardwareProvider's* customers in most cases focus on the underlying hardware technology. At this stage, the software platform is only seen as a complementary service to implement specific, vertical end-to-end use cases. Generic platform features like device management, user management, or generic standard interface implementations are used rarely within these proof-of-concepts. After a successful proof-of-concept, the customer decides to stay with the technology in form of a permanent hardware installation. Although this entails the installation of more hardware components in many cases, the scope of software features being used remains mostly the same. However, ensuing customer workshops often result in the identification of further projects. These projects consider not only the implementation of the similar use cases at other sites. Instead, identified use cases increasingly exploit platform characteristics of *HardwareProviders* software components. This is, customers' requirements develop more towards use cases utilizing the platform's standard interfaces to connect external data sources, for integration with existing systems or to connect with other external services on top. Furthermore, cross-cutting use cases depend on the scalability of a platform's architecture.

Benefits: Following such an evolvable customer journey approach has several benefits for *HardwareProvider*. First, it allows the company to exploit its innovative hardware technology and existent expertise by using it as an entry point for launching a more sustainable platform business-model at the customer. Furthermore, this step-by-step approach is the foundation for involved stakeholders at the customer site to develop a deeper understanding concerning the technology and its potential. By also building trust in *HardwareProvider* as a technology vendor, these stakeholders serve as internal promoters for follow-up projects. In general, customer investments in platform adoption and training for software usage, as well as already implemented use cases, including customization and integration with other IT systems, entail a certain lock-in effect. This was also summarized by one interview partner: '*And once [...] you anchored yourself horizontally with other systems, then it is very difficult to lose that position again. This way, you can establish yourself very well and thereby make yourself interesting for follow-up business [...].'* [IP5]

Challenges: In the first place, small proof-of-concept projects, including continuous consulting services and close collaboration, are not profitable. This evolving approach results in long sales cycles requiring *HardwareProvider* to make big investments in the technology not being covered by current revenues.

From a technical perspective, the chosen approach cannot be realized as seamlessly as it may be assumed. This is, further use cases are not just implemented by activating additional features on the platform. For example, an increased scalability required by later use case extensions (e.g., additional data sources) causes changes in the infrastructure. An interview partner outlines the reason for having two different versions: *'For this scalability, which is that this cluster is functioning, it requires a minimum of three servers. Without that, it doesn't make sense. And not every customer wants to put up three servers, if it is not needed. Thus, we also have another mode, which operates on only one server.'* [IP4] Further technical challenges include version management and security aspects related to network integration.

Lastly, *HardwareProvider* faces challenges concerning pricing. At first glance, a mixed offering (hardware plus software) may allow for more flexibility in cost allocation. However, in reality, the combination of one-time hardware costs and recurring software costs associated with an uncertainty about future software development and usage poses a problem. One strategy to solve this problem is the modularization of platform features. Thereby, value-based vs. cost-based pricing considerations as well as questions about feature granularity (e.g., statistic module vs. single charts) and licenses control mechanisms arise.

5 Discussion

Findings of the conducted case study shed light on a platform owner's challenging situation during early platform development. While the previous chapter explained our observations in more detail, we now discuss findings in the context of extant platform literature. We find that the emergence of the identified trade-off, which was rarely addressed in the literature so far, is attributable to characteristics specific to the B2B market. Especially, the complexity of B2B solutions, length of sales processes, bargaining power of customers, as well as the market fragmentation resulting in a lack of standardization, causes the differences to findings of existing literature, which are mostly based on examples from the B2C context [12, 13, 14, 15, 49].

Results of this case study further show, that, initially, customers in this market are mostly interested in single vertical end-to-end solutions for certain use cases. Thus, indirect network effects as a central and valuable mechanism of platforms do not occur because customers do not value the availability of other use cases on the platform. Consequently, the well-researched chicken-and-egg problem is less relevant in this context [32]. Nevertheless, customers expect a flexible and open solution to be able to develop further applications in their own IT departments or jointly with other third-party developers. Hence, platform-specific design principles (modularity, stable interfaces) and provision of boundary resources (documentation, developer communities etc.) are still relevant aspects to consider for the case company [14, 31].

Analyzing the case company's situation and chosen strategy, we provide insights to the prominent research question examining a company's decision to follow a platform strategy in the complex B2B market, considering differences between a horizontal and a vertical strategy [4, 21, 22]. In particular, we outlined the need for a hybrid strategy caused by anticipated consequences for internal and external stakeholder. Furthermore, this work also contributes to theory dealing with 'make-or-join' decisions being made by companies in emerging platform markets, that is, the decision whether to implement a new platform or to join an existing platform ecosystem [50]. Due to a complex and vast market situation, the immature state (in terms of technology and customer base) of existing platforms in the market, as well as the inability of these solutions to serve specific technical requirements (e.g. real-time and scalability), the case company decided to pursue a 'make' strategy.

This paper also adds to literature researching platform development from a market perspective. Unlike the assumption of a homogenous shift towards a horizontal market architecture [27], the results of this case study reveal, that in B2B markets, this transition from vertical integrated products towards a horizontally layered architecture can happen for every customer individually on different points in time. As a result, a company striving for platform ownership needs to provide a hybrid solution first. By describing *HardwareProvider's* hybrid strategy utilizing an evolvable customer journey approach, this paper presents a possible solution for this situation. We also outline the benefit of this strategy as yielding strong lock-in effects through investments by the customers for user education and technical system integration. Due to the complexity of B2B products, these investments are bigger, switching costs are higher and lock-in effects are stronger compared to those of B2C platforms. Thus, we consider it as unlikely, that the prominent ‘Winner takes all’ mechanisms [32] will apply in this context. Rather we expect the market to be composed of several niche platforms being linked with each other.

Our study also contributes to practice: By breaking down the current situation of a real company in this field, our work reveals important aspects and challenges to consider when working on a business model in an emerging B2B platform market. However, practitioners should consider, that results presented here are derived from a case study at a start-up and thus may not apply to established companies to the same extent [cf. 50, 51].

6 Conclusion

The purpose of our exploratory case study was to deepen the understanding of prospective platform owners’ situation in the context of Industry 4.0. Thereby, we derived novel insights about the early stage of platform development and the challenging trade-off between following a vertical product strategy or the implementation of a horizontal platform business model. We not only presented the identified circumstances fueling this challenging situation, but also outlined the platform owner’s approach of pursuing an evolvable customer journey as one strategy to counteract this trade-off. Although this strategy yields strong benefits like customer lock-in, it still includes business and product related challenges.

However, our study is also subject to several *limitations* laying the foundation for *future work*. Conducting a single case study allowed us to gain in-depth understanding of the company’s situation in the complex emerging IIoT market. Nevertheless, the generalization of findings is restricted. Furthermore, we want to point out, that the long-term success of the presented strategy undertaken by the case company could not be evaluated, as the platform is still in a very early stage. To overcome these limitations, we suggest multiple, long-term case studies also addressing differences in platform type (e.g., edge vs. cloud or platform for equipment supplier vs. for equipment operator) and platform owner’s characteristics (background, size, age) while evaluating findings. Lastly, presented results focus on the technical development and design of the platform. Besides that, building an ecosystem is a key aspect of successful platforms. While the onboarding process of customers has been addressed in more detail, onboarding of the supply side was not in the focus of this work. However, in the course of this exploratory case study we came across interesting aspects differing from extant literature (e.g., partner acquisition, motivation, collaboration). Thus, we call for future work further investigating partner management strategies in the context of IIoT platforms and hope that our work can provide a helpful starting point.

7 Acknowledgements

We thank the German Federal Ministry for Economic Affairs and Energy for funding this research as part of the project 01MD15001D (ExCELL).

References

1. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative Industrie 4.0. https://www.acatech.de/Publikation/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/ (2013)
2. Schreier, J.: Title. Industry of Things (2018), https://www.industry-of-things.de/schneller-iiot-einstieg-plattformunabhaengigkeit-entscheidet-a-718049/
3. Schreieck, M., Hakes, C., Wiesche, M., Krcmar, H.: Governing Platforms in the Internet of Things. International Conference on Software Business, Essen, Germany (2017)
4. Tiwana, A.: Platform Ecosystems: Aligning Architecture, Governance, and Strategy. Morgan Kaufmann, Burlington, MA (2014)
5. Bakos, Y., Katsamakas, E.: Design and Ownership of Two-Sided Networks: Implications for Internet Platforms. Journal of Management Information Systems 25(2), 171-202 (2008)
6. Lin, M., Li, S., Whinston, A.B.: Innovation and Price Competition in a Two-Sided Market. Journal of Management Information Systems 28(2), 171-202 (2011)
7. Koh, T.K., Fichman, M.: Multihoming Users' Preferences for Two-Sided Exchange Networks. MIS Quarterly 38(4), 977-996 (2014)
8. Baldwin, C.Y., Woodard, C.J.: The Architecture of Platforms: A Unified View. In: Gawer, A. (ed.) Platforms, Markets and Innovation, pp. 19-44. Edward Elgar, Cheltenham, UK (2009)
9. Eisenmann, T.R., Parker, G., Van Alstyne, M.: Opening Platforms: How, When and Why? In: Gawer, A. (ed.) Platforms, Markets and Innovation, pp. 131-162. Edward Elgar, Cheltenham, UK (2009)
10. Boudreau, K.J.: Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. Organization Science 23(5), 1409-1427 (2012)
11. Schreieck, M., Wiesche, M., Krcmar, H.: The Platform Owner's Challenge to Capture Value–Insights from a Business-to-Business IT Platform. Thirty Eighth International Conference on Information Systems, Seoul, South Korea (2017)
12. Song, P., Xue, L., Rai, A., Zhang, C.: The Ecosystem of Software Platform: A Study of Asymmetric Cross-Side Network Effects and Platform Governance. MIS Quarterly 42(1), 121-142 (2018)
13. Boudreau, K.J.: Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. Management Science 56(10), 1849-1872 (2010)
14. Ghazawneh, A., Henfridsson, O.: Balancing platform control and external contribution in third-party development: the boundary resources model. Information Systems Journal 23(2), 173-192 (2013)
15. Ye, H., Kankanhalli, A.: User Service Innovation on Mobile Phone Platforms: Investigating Impacts of Lead Userness, Toolkit Support, and Design Autonomy. MIS Quarterly 42(1), 165-187 (2018)
16. Schreieck, M., Wiesche, M.: How Established Companies Leverage IT Platforms for Value Co-Creation – Insights from Banking. Twenty-Fifth European Conference on Information Systems, Guimarães, Portugal (2017)
17. Evans, D.S., Hagiu, A., Schmalensee, R.: Invisible Engines: How Software Platforms Drive Innovation and Transform Industries. MIT Press, Cambridge, MA (2006)
18. Iyer, B., Lee, C.-H., Venkatramen, N., Vesset, D.: Monitoring Platform Emergence: Guidelines from Software Networks. Communications of the Association for Information Systems 19(1), 1-13 (2007)

19. Eisenmann, T., Parker, G., Van Alstyne, M.: Platform Envelopment. Strategic Management Journal 32(12), 1270–1285 (2011)
20. Gawer, A.: Platform Dynamics and Strategies: From Products to Services. In: Gawer, A. (ed.) Platforms, Markets and Innovation, pp. 45-76. Edward Elgar, Cheltenham, UK (2009)
21. Gawer, A., Cusumano, M.A.: How Companies Become Platform Leaders. MIT Sloan Management Review 49(2), 28-35 (2008)
22. Parker, G., Van Alstyne, M., Jiang, X.: Platform Ecosystems: How Developers Invert the Firm. MIS Quarterly 41(1), 255-266 (2017)
23. Eisenhardt, K.M.: Building Theories from Case Study Research. Academy of Management Review 14(4), 532-550 (1989)
24. Yin, R.K.: Case Study Research: Design and Methods. Sage, Los Angeles, CA (2009)
25. Van de Ven, A.H.: Engaged Scholarship: A Guide for Organizational and Social Research. Oxford University Press, New York (2007)
26. Schreieck, M., Wiesche, M., Krcmar, H.: Design and Governance of Platform Ecosystems-Key Concepts and Issues for Future Research. Twenty-Fourth European Conference on Information Systems, Istanbul, Turkey (2016)
27. Yoo, Y., Henfridsson, O., Lyytinen, K.: Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. Information Systems Research 21(4), 724-735 (2010)
28. Porter, M.E., Heppelmann, J.E.: How Smart, Connected Products Are Transforming Competition. Harvard Business Review 92(11), 64-88 (2014)
29. Katz, M.L., Shapiro, C.: Systems Competition and Network Effects. Journal of Economic Perspectives 8(2), 93-115 (1994)
30. Iansiti, M., Zhu, F.: Dynamics of Platform Competition: Exploring the Role of Installed Base, Platform Quality and Consumer Expectations. Twenty Eighth International Conference on Information Systems, Montreal (2007)
31. Tiwana, A.: Evolutionary Competition in Platform Ecosystems. Information Systems Research 26(2), 266-281 (2015)
32. Eisenmann, T., Parker, G., Van Alstyne, M.W.: Strategies for Two-Sided Markets. Harvard Business Review 84(10), 92-101 (2006)
33. Adner, R., Kapoor, R.: Value Creation in Innovation Ecosystems: How the Structure of Technological Interdependence affects Firm Performance in New Technology Generations. Strategic Management Journal 31(3), 306-333 (2010)
34. Dhanaraj, C., Parkhe, A.: Orchestrating Innovation Networks. Academy of Management Review 31(3), 659-669 (2006)
35. Scholten, S., Scholten, U.: Platform-based Innovation Management: Directing External Innovational Efforts in Platform Ecosystems. Journal of the Knowledge Economy 3(2), 164-184 (2012)
36. Smedlund, A., Faghankhani, H.: Platform Orchestration for Efficiency, Development and Innovation. 48th Hawaii International Conference on System Sciences. IEEE Computer Society, Hawaii (2015)
37. Rickmann, T., Wenzel, S., Fischbach, K.: Software Ecosystem Orchestration: The Perspective of Complementors. Twentieth Americas Conference on Information Systems, Savannah (2014)
38. Eaton, B., Elaluf-Calderwood, S., Sorensen, C., Yoo, Y.: Distributed Tuning of Boundary Resources: the Case of Apple's iOS Service System. MIS Quarterly 39(1), 217-243 (2015)
39. Schreieck, M., Finke, R., Wiesche, M., Krcmar, H.: Sandbox vs. Toolbox – Analyzing Boundary Resource in B2B Software Platforms. European Workshop on Software Ecosystems, Darmstadt (2017)

40. Cusumano, M.A., Gawer, A.: The Elements of Platform Leadership. MIT Sloan Management Review 43(3), 51-58 (2002)
41. Zhu, F., Furr, N.: Products to Platforms: Making the Leap. Harvard Business Review 94(4), 72-78 (2016)
42. Urquhart, C.: Grounded Theory for Qualitative Research - A Practical Guide. Sage, London (2013)
43. Bowen, P., Goel, A., Schallehn, M., Schertler, M.: Choosing the Right Platform for the Industrial IoT. https://www.bain.com/insights/choosing-the-right-platform-for-the-industrial-iot/ (2017)
44. Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. VS Verlag für Sozialwissenschaften, Wiesbaden (2010)
45. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine, Chicago, IL (1967)
46. Wiesche, M., Jurisch, M.C., Yetton, P., Krcmar, H.: Grounded Theory Methodology in Information Systems Research. MIS Quarterly 41(3), 685-701 (2017)
47. Glaser, B.G.: Theoretical Sensitivity: Advances in the Methodology of Grounded Theory. Sociology Press, Mill Valley, CA (1978)
48. Lemon, K.N., Verhoef, P.C.: Understanding Customer Experience Throughout the Customer Journey. Journal of Marketing 80(6), 69-96 (2016)
49. Hilkert, D., Benlian, A., Hess, T.: Motivational Drivers to Develop Apps for Social Software-Platforms: The Example of Facebook. Proceedings of the Sixteenth Americas Conference on Information Systems, Lima, Peru (2010)
50. Schreieck, M., Wiesche, M., Krcmar, H.: Multi-Layer Governance in Platform Ecosystems of Established Companies. Academy of Management Proceedings, Chicago, Illinois (2018)
51. Weiß, N., Schreieck, M., Brandt, L.S., Wiesche, M., Krcmar, H.: Digitale Plattformen in der Automobilbranche – Herausforderungen und Handlungsempfehlungen. HMD Praxis der Wirtschaftsinformatik 55(2), 349-365 (2018)

Machine Learning und Complex Event Processing: Effiziente Echtzeitauswertung am Beispiel Smart Factory

Jonas Wanner[1], Christopher Wissuchek[1], Christian Janiesch[1]

[1] Julius-Maximilians-Universität Würzburg, Würzburg, Germany
{jonas.wanner, christian.janiesch}@uni-wuerzburg.de,
christopher.wissuchek@gmail.com

Abstract. Durch die Verbindung zwischen physischen Maschinenteilen und digitalen Services werden mit Cyber-physischen Systemen in Smart Factories viele datenbasierte Optimierungen möglich. Ein wichtiger Bestandteil dieser sogenannten Smart Factories kann die Technologie Complex Event Processing (CEP) sein. CEP erlaubt Echtzeitauswertungen komplexer Events, i. S. v. kombinierten Datenwerten aus unterschiedlichen Quellen. Damit können u. a. anomale Prozessabläufe identifiziert und lokalisiert werden. Eine aktuelle Beschränkung der Wirkungsfähigkeit ist die hauptsächlich deklarative und reaktive Implementierung von CEP. Eine Erweiterung um Ansätze aus dem Machine Learning (ML) ist daher vielversprechend. Es fehlt jedoch an einer aktuellen Übersicht zu Verbindungen von CEP und ML innerhalb der Forschung sowie deren Transferfähigkeit auf Smart Factories. Unser Beitrag liefert (1) eine Synthese der bislang erforschten CEP-ML-Kombinationen, wobei sich Supervised Learning als überwiegender Kombinationsansatz zeigt, und (2) eine Übertragung der Potenziale für die Verwendung in Smart Factories. Hier zeigten sich reaktive Maßnahmen als bisheriger Forschungsschwerpunkt.

Keywords: Machine Learning; Complex Event Processing; Real-time Event Processing; Smart Factory; Literature Review

1 Einleitung

Die nutzbare Datenmenge in der modernen Industrie steigt stetig an [1]. Inzwischen beschränkt sich die Nutzung der Analysepotenziale nicht nur auf historische Datenbestände, sondern verstärkt auf die Auswertung von Echtzeitdaten. Dies erlaubt eine situationsabhängige und dynamische Entscheidungsfindung [2]. Ein Realisierungsansatz dafür kann Complex Event Processing (CEP) sein, ein Technologie-Paradigma zur Verarbeitung kontinuierlicher Datenströme auf Basis von Ereignissen [3].

Gerade in Smart Factories lassen sich mit dieser Technologie Einsparungen durch die Echtzeitauswertung von Sensordaten erzielen [4]. Trotz der großen Potenziale von CEP existiert jedoch eine fundamentale Einschränkung: CEP verarbeitet und analysiert Ereignisse auf Basis manuell vorspezifizierter, deklarativer Regeln [5, 6]. Durch die regelbasierte Umsetzung [6] ist eine Proaktivität schwierig umsetzbar. Ebenso erfordert

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

der reaktive Charakter teure manuelle Identifikations- und Implementierungsprozesse [7].

Forscher aus der Domäne des CEP greifen infolgedessen zuletzt verstärkt auf Machine Learning (ML) zurück, um die (teure) deklarative Natur von CEP zu verändern und zu verbessern [8-12]. ML ist ein Ansatz, in dem Computersysteme automatisiert auf Basis von Algorithmen aus Erfahrung lernen [13].

Zusammenfassend sollen die aktuellen CEP-Restriktionen durch geeignete ML-Möglichkeiten aufgeweicht und wenn möglich ganz gegenstandslos werden. Der Fokus dieser Arbeit liegt dabei auf der Anwendbarkeit in Cyber-physischen Systemen moderner Smart Factories. Es lassen sich damit zwei Forschungsfragen ableiten:

(1) *Welche ML-Verfahren werden in Kombination mit CEP eingesetzt und wie sind diese zu systematisieren?*
(2) *Welche Potenziale leiten sich aus dem Einsatz von ML im CEP von Industriefertigungsanlagen in Smart Factories ab?*

Zur Beantwortung der Forschungsfragen ist dieser Beitrag wie folgt aufgebaut: In Kapitel 2 wird ein Überblick zu den theoretischen Grundlagen von Smart Factories, CEP und ML gegeben. Kapitel 3 erläutert die angewendete Forschungsmethodik. Kapitel 4 beantwortet Forschungsfrage (1) auf Basis einer literaturgestützten Taxonomie zur Lokalisierung potenzieller Anwendungsmöglichkeiten. Die erlangten Erkenntnisse werden anschließend innerhalb von Kapitel 5 zur Beantwortung von Forschungsfrage (2) auf Herausforderungen und Verbesserungspotenziale für Anlagen in Smart Factories übertragen. In Kapitel 6 erfolgt eine kritische Stellungnahme zu den erarbeiteten Erkenntnissen. Ebenso werden erkannte Forschungslücken aufgezeigt. Die Arbeit schließt mit einem Fazit.

2 Smart Factories, CEP und ML

Cyber-physische Systeme und Smart Factories. Die zentralen Produktionsfaktoren der Industrie 4.0 sind Informations- und Kommunikationstechnologien [14]. Sie werden dazu genutzt Cyber-physische Systeme (CPS) zu entwickeln. Dabei handelt es sich u. a. um Objekte, Gebäude oder Produktionsanlagen, die durch eingebettete Systeme kommunikationsfähig werden und ihre Umgebung mittels verbundener Sensorik erfassen können. Sie sind vernetzbar und offenbaren über echtzeitfähige Messungen von Ereignissen (sog. Events) neue Optimierungs- und Automatisierungspotenziale [14].

Complex Event Processing. Eine Möglichkeit der Nutzenmachung dieser Events ist CEP, ein Technologie-Paradigma zur Echtzeitverarbeitung kontinuierlicher Ereignis-Datenströme [3]. Die CEP-Referenz-Architektur besteht aus drei Komponenten. Der (1) *Event Producer* überträgt die gesammelten Events an den (2) *Event Processor*, die zentrale Verarbeitungseinheit. Dieser verarbeitet Events regelbasierend auf Basis manuell und deklarativ zu definierender Agenten zur Erkennung anomalen Verhaltens [7]. Die Verbesserung der Fähigkeiten dieser sogenannten *Event Processing Agents* durch ML ist der Fokus des vorliegenden Beitrags. Nach der Verarbeitung werden die Ergebnisse an festgelegte (3) *Event Consumer* übergeben, welche die Ereignisbehandlung übernehmen [2, 15].

ML-Verfahren. Die bisherigen CEP-Einschränkungen sollen durch Möglichkeiten des ML verbessert werden. Dies erlaubt die automatisierte CEP-Regelbildung für bekannte und ggf. unbekannte Anomalien. Zum Beispiel könnte eine zu definierende, Schwellenwert-basierte Filterung durch eine ML-trainierte Klassifikation ersetzt werden. ML bietet dazu unterschiedliche Verfahren zur Schulung von Algorithmen. *Supervised Learning* umschreibt einen Lernansatz, bei welchem Inputs bewusst vordefinierten Outputs zugeordnet werden. Durch die Übergabe der Input-/ Output-Paare zu Trainingszwecken kann die Maschine künftig neue Fälle automatisiert zuordnen [16-18]. Diese können weiter in Klassifizierungs- und Regressionsansätze unterkategorisiert werden. Beim *Unsupervised Learning* wird auf die Übergabe der Input-/ Output-Paare verzichtet. Der verwendete Algorithmus erhält ausschließlich Inputs. Als Ziel gilt das Erkennen bisher unentdeckter Muster [17, 19, 20]. Hier wird i. d. R. mit Clustering und Dimensionsreduktion gearbeitet. Das *Reinforcement Learning* ist ein weiterer Lernansatz zum Verstehen und zur Automatisierung von zielgerichtetem Lernen und Entscheidungsfindungsprozessen [21]. Durch Versuch und Irrtum in Kombination mit Belohnungen wird ein Lernagent geschult, sodass er je nach angenommenen Zuständen und Aktionen Belohnungen maximiert [18, 22, 23].

3 Methodisches Vorgehen

Zur Beantwortung der Forschungsfragen wurde das Rahmenwerk zur strukturierten Literaturrecherche nach vom Brocke et al. [24] verwendet.

Rechercheumfang. Der festzulegende Rechercheumfang basiert auf der Taxonomie nach Cooper [25]. Als Fokus gelten *Forschungsergebnisse* und *Anwendungen*. Das Ziel ist die *Integration* der Forschungsbereiche ML und CEP. Es wird eine *neutrale Darstellung* mit *repräsentativer Abdeckung* gewählt. Die Ergebnisse selbst werden *konzeptionell* zur Beantwortung der Forschungsfragen organisiert. Als Zielgruppe haben wir *spezialisierte Forscher* aus den Bereichen von ML und oder CEP sowie *allgemeine Forscher* aus der Wirtschaftsinformatik bestimmt.

Konzeptualisierung. Das Suchergebnis muss ein Verfahren aus der Domäne des ML im CEP einsetzen und auf Anwendungen in Smart Factories übertragbar sein.

Literatursuche. Bei der *Datenbankauswahl* wurden sieben Datenbanken aus den Bereichen der Informatik und Betriebswirtschaftslehre gewählt (vgl. *Abbildung 1*). Für die *Keyword-Suche* wurden gängige Begriffe aus dem CEP genutzt (Kurzform: *Event & Processing | -driven*) und mit dem engl. Oberbegriff „Machine Learning“ verbunden.[1]

[1] Eine Übersicht der definierten Suchterme je wissenschaftlicher Datenbank sowie die vollständigen Referenzen zu Tabelle 1 können im digitalen Anhang abgerufen werden unter: https://bit.ly/2qSdhP8.

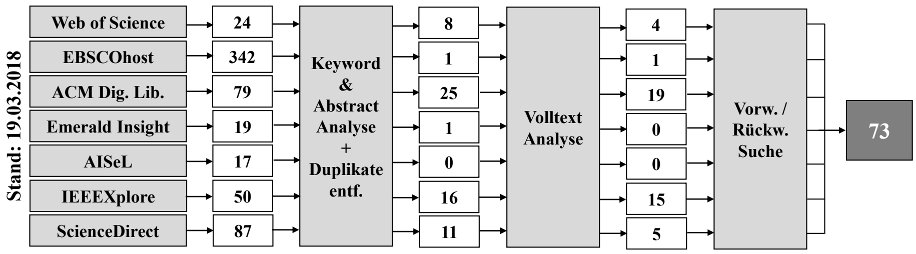


Abbildung 1: Prozess der Literatursuche

Es wurden im ersten Schritt 618 potentiell relevante Werke identifiziert. Nach der Analyse aller Werke bzgl. Titel, Abstract, Keywords und Duplikaten konnten 62 relevante Publikationen ermittelt werden. Die Volltextanalyse reduzierte die Anzahl auf 44. Ein Gros an Suchergebnisse fand sich in informatiknahen Datenbanken (*IEEEXplore* und *ACM Digital Library*). Letztere publiziert u. a. Beiträge der für das Thema hoch relevanten Konferenz Distributed Event-Based Systems (DEBS). Die *Rückwärtssuche* wurde auf Basis der extrahierten Quellen aus den Werken durchgeführt, die *Vorwärtssuche* über Google-Scholar-Zitationsdaten. Dies führte zu weiteren 29 Ergebnissen, sodass insgesamt 73 Beiträge gefunden werden konnten.

4 Auswertung der Literaturanalyse

Ergebnis der Synthese. Die Synthese (vgl. *Abbildung 2*) zeigt, dass das Lernparadigma *Supervised Learning* im CEP dominiert (n=46). Der Großteil der Arbeiten behandelt dabei den algorithmischen Ansatz der *Klassifizierung* (n=23).

Das Lernparadigma *Unsupervised Learning* findet hingegen weniger Anwendung (n=6). *Clustering* zeigt sich hier als dominanter Ansatz (n=5), wobei ein Beitrag einen *dimensionsreduktionsbasierten Ansatz* verfolgt.

Beliebter ist die Kombination (Englisch: *Combined Learning*) aus *Unsupervised* und *Supervised Learning* (n=13). Die meisten Autoren (n=12) nutzen hierbei eine Verbindung aus *Clustering* und *Klassifikation*. Ein Ansatz behandelt *Clustering* erweitert um ein *probabilistisches Modell* (n=1).

Reinforcement Learning für CEP wurde bisher nicht beachtet und scheint damit potentieller möglicher Gegenstand weiterer Forschung. Weitere alternative Ansätze, welche sich nicht in das gewählte Klassifikationsschema einordnen ließen, wurden unter *Weitere Beiträge* (n=8) aufgelistet. Beispiele finden sich mit Algorithmen aus der Data Mining Domäne [26] oder Eigenentwicklungen wie bei Margara et al. [27] mit deren Windows Learner.

Zeitliche Entwicklung. Die Forschungsdomäne kann als jung bezeichnet werden mit dem frühsten Beitrag aus dem Jahr 2007 [28]. Über die Folgejahre zeigt sich ein steigendes Forschungsinteresse mit Höhepunkten in den vergangenen Jahren 2016 (n=18) und 2017 (n=17). Das Jahr 2018 ist nicht vollständig erfasst. Diese Tendenz verdeutlicht die Aktualität und Themenrelevanz sowie die verschiedenartigen Anwendungsmöglichkeiten des ML im CEP.

Auf Ebene der Lernparadigmen ist festzustellen, dass anfangs vor allem *Supervised Learning* betrachtet wurde. Hierbei lag der Fokus zunächst auf Klassifizierungsansätzen. Ab dem Jahr 2013 findet eine Diversifikation statt. Weitere Ansätze, wie bspw. regressionsbasierte Verfahren, finden vermehrt Betrachtung. Dennoch bleiben Klassifikationsverfahren durchgehend dominant. *Unsupervised Learning* ist in der CEP-Forschung hingegen (bisher) unterrepräsentiert. Die erste Publikation aus dem Jahr 2010 verfolgt einen dimensionsreduktionsbasierten Ansatz. Ab 2013 werden vereinzelt Clustering-basierte Beiträge veröffentlicht, wobei das Maximum im Jahr 2017 (n=2) liegt. Ein stärkeres Interesse zeigt sich bei der *Kombination* beider Paradigmen. Nach nur einer Publikation im Jahr 2014, finden sich 2016 fünf und 2017 sieben Beiträge.

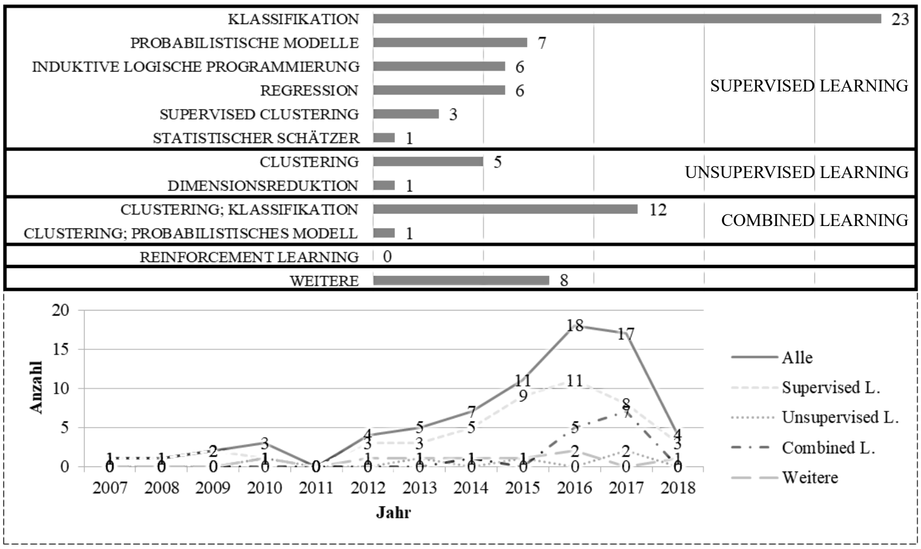


Abbildung 2: Zuordnung der Beiträge zu den ML-Lernparadigmen

Fazit. Im CEP findet überwiegend *Supervised Learning* Anwendung. Dies lässt sich aufgrund des reaktiven Charakters in Kombination mit historischen Eventdaten erklären. *Unsupervised Learning* wird dementsprechend meist in *Kombination* eingesetzt. Damit lassen sich Vorteile aus beiden Paradigmen nutzen, wobei hierbei weiterer Forschungsbedarf besteht. Eine fundierte Eignungsprüfung des Ansatzes *Reinforcement Learning* für CEP fehlt hingegen. Es zeigt sich aus der historischen Analyse, dass die Forschung im Bereich CEP und ML an Wirkungskraft gewinnt und nicht nur punktuell einzelne Verfahren betrachtet. Es ist vielmehr ein Trend zu erkennen, dass verschiedene Verfahren kombiniert werden, um einen erkennbaren Mehrwert in der Echtzeitverarbeitung von Events zu schaffen.

5 Anwendungspotenziale in Smart Factories

5.1 Zielbereiche von Produktionsanlagen

Aus betriebswirtschaftlicher Perspektive bestehen in Produktionsanlagen vier Zielbereiche [29]: Kosten, Qualität, Zeit und Flexibilität. Unter *Kosten* wird mehrheitlich die bestmögliche Allokation der eigenen Ressourcen verstanden. *Qualität* umschreibt die Annahme und Schätzung einer Übereinkunft des Produkts mit den eigenen Zielvorgaben basierend auf dem Produktionsprozess. *Zeit* definiert sich als Quelle eines Wettbewerbsvorteils i. S. v. Ansätzen wie Just-in-time und als fundamentale Messinstanz für die Produktionsperformance. *Flexibilität* beinhaltet die Fähigkeit eines Industrieherstellers, sein Produktionssystem effizient zu verändern – sowohl hinsichtlich der Produktvielfalt als auch der Herstellmenge [30].

Dieselben Zielbereiche bestehen auch in Smart Factories [4]. Durch den Einsatz von CPS und der verbundenen Infrastruktur sollen potenzielle Verbesserungen in allen vier Bereichen realisierbar sein [4]. Um Potenziale durch die Kombination aus CEP und ML aufzuzeigen, wird im nachfolgenden Kapitel von derartigen Industrieanlagen ausgegangen.

5.2 Synthese potenzieller Lösungsansätze durch CEP und ML

Synthese. Auf Basis der Rechercheergebnisse lassen sich sowohl reaktive als auch proaktive Einsatzpotenziale für Smart Factories erkennen. Wie in *Tabelle 1* ersichtlich überwiegen Forschungsbemühungen zu reaktiven Möglichkeiten. Ebenso sind die in *Abbildung 2* unter *Weitere* aufgeführten Beiträge ausschließlich auf reaktive Maßnahmen gerichtet. Aus Platzgründen wurde jedoch auf die Aufnahme in *Tabelle 1* verzichtet, zumal es sich weitestgehend um Spezialfälle handelt oder kein spezifischer Algorithmus genannt wird. Ein Hinweis zu weiteren Details findet sich in Fußnote 1.

Reaktive Maßnahmen. Reaktive Maßnahmen von CEP in Kombination mit ML sollen effizientere Abläufe für Fehler- bzw. Schadensfälle ermöglichen. Dies erlaubt die Aufbereitung wichtiger Informationen für die Fehlerlokalisierung und -behebung. Bei Beginn des CEP-Monitorings in Industrieanlagen sind allerdings i. d. R. nicht alle Störfälle bekannt. Eine Fortschreibung der Mustererkennung ist daher essentiell. Dies erfordert eine manuelle Identifizierung neuer Anomalien. Die Aufgabe impliziert hohes technisches Wissen, bei vertiefender Domänenkenntnis [5, 6]. Eine Unterstützung kann ML bieten. Damit werden im Idealfall bisher unbekannte Eventmuster vollautomatisch aufdeckbar und für die Zukunft aufbereitet [28, 59].

Einen direkten Anwendungsbezug in Smart Factories bietet der Ansatz von Metz et al. [35]. Sie nutzen Produktionsdaten in Kombination mit Prozessdaten, um mit Hilfe einer ML-Komponente neue Regeln über Entscheidungsbäume zu klassifizieren und diese für die CEP-Einheit zu nutzen. Ebenso findet sich der Einsatz von Kamerasystemen und bildverarbeitenden Sensoren in Produktionsprozessen [42]. Diese bieten wertvolle Informationen für Smart Factories, um eine präzisere Mustererkennung und -adaption zu realisieren.

Tabelle 1: Potenziale beim Einsatz von CEP und ML in Smart Factories (für die vollständige Einordnung aller analysierten Quellen vgl. Fußnote 1)

Lernparadigma	ML-Ansatz	Algorithmus	Reaktive Maßnahmen (∑=39) ∑	Beispiel	Proaktive Maßnahmen (∑=35) ∑	Beispiel
Supervised Learning	Klassifikation	Support Vector Machine	4	[31]	1	[32]
		Shapelet-basierte Algorithmen	2	[33]	3	[34]
		Entscheidungsbäume	2	[35]	2	[36]
		Künstlich neuronale Netze	4	[12]	1	[37]
		Sliding Window, Aggressive and Pessimist Classifier	1	[38]	0	
		Markov Model	1	[7]	1	[39]
		Diskriminanzanalyse	3	[12]	0	
		K-nearest Neighbors	1	[31]	0	
		Naive Bayes	1	[31]	1	[36]
		FURIA	3	[40]	0	
	Probabilistisches Modell	Bayessche Netzwerke	0		6	[36]
		Markov Model	1	[41]	0	
		Conditional Density Estimation	0		1	[42]
	Induktive logische Program.	Eigenentwicklung	6	[43]	0	
	Regression	Support Vector Regression	0		3	[44]
		Symbolic Regression	0		1	[44]
		Adaptive Moving Window Regression	0		2	[45]
	Supervised Clustering	k-Means Adaption	0		1	[46]
		Eigenentwicklung	1	[47]	1	[48]
	Statistischer Schätzer	Discrete Kalman Filter	1	[6]	0	
Unsupervised Learning	Clustering	Density-based Clustering	2	[49]	0	
		Complete-link and Single-link Clustering	1	[50]	0	
		Fixed-width Clustering	1	[51]	0	
		k-Means	1	[5]	0	
	Dimensionsreduktion	Principal Component Analysis	1	[52]	0	
Combined Learning	Clustering; Klassifikation	k-Means; Markov Model	0		8	[53]
		k-Means; Gaussian Mixture Model	1	[54]	0	
		Hierarchical Clustering; Entscheidungsbäume; k-Means; Naive Bayes, SVM	0		1	[55]
		Unspezifizierter Clustering-Algorithmus; Naive Bayes	1	[56]	0	
	Clustering; Probabilistisches Modell; Topic Mining	Density-based Clustering; Bag-of-words; Markov Model	0		1	[57]
	Clustering; Klassifikation; Kollaboratives Filtern	k-Means; Complementary Naive Bayes; ALSWR	0		1	[58]

Darüber hinaus werden Sensornetze in Kombination mit Klassifizierungsalgorithmen genutzt [38]. Die Erkennung anomalen Verhaltens erfolgt dabei über Analysen von Events aus Druck-, Vibrations-, Temperatur- und Näherungssensoren [60]. Auch Ansätze aus weiteren Forschungsbereichen zeigen Potenzial für die Problemlösung auf. Einige der Autoren [41, 43] setzen auf Videoüberwachungs-Datensätze, um Aktivitäts- bzw. Eventmuster zu erlernen. Hierbei wird bis auf eine Ausnahme einer Markov-Model-Verwendung in [41] auf Induktive Logische Programmierung gesetzt. Die Anwendung erfolgt allerdings außerhalb von Smart Factories, ist aber potenziell übertragbar.

Weitere Autoren [40, 56] nutzen Beschleunigungssensor-basierte Ereignisströme, um automatisiert Bewegungsmuster von Personen zu erkennen. Für das ML wird bei [40] eine Klassifizierung mittels des Fuzzy Unordered Rule Induction Algorithm (FURIA) verwendet. [56] nutzen hingegen eine Kombination aus Clustering und Klassifikation auf Basis von Naive Bayes und eines unspezifizierten Clustering-Ansatzes. In produktionsspezifischen Szenarien bietet dies jeweils u. a. die Möglichkeit zur Überwachung und Steuerung von industriellen Roboterarmen [61]. Eine manuelle Implementierung der Bewegungsprofile als Events in einem CEP-System wäre dagegen mit großem zeitlichem Aufwand und hoher technischer Expertise (Kosten) verbunden.

Ein weiteres Anwendungsfeld ist die Netzwerksicherheit. Durch den Einsatz von CPS werden neue Cyber-Angriffsflächen eröffnet, welche eine Betriebsstörung oder Produktionsausfälle verursachen können [62]. Zur Prävention resultierender Zeitverluste und Kosten setzen [6, 26] auf CEP-basierte Intrusion Detection Systeme, welche um ML-Komponenten erweitert werden. [6] nutzen hierfür Discrete Kalman Filters, wohingegen [26] eine Kombination aus Markov Logic und Data Mining Ansätzen einsetzen. In beiden Fällen wird eine dynamische und adaptive Anpassung der Angriffsmuster möglich ohne die Notwendigkeit von manuellen Eingriffen und wiederkehrender, aufwändiger Nacharbeitung.

Durch die aufgezeigten Maßnahmen werden primär die manuelle Überwachungstätigkeit und die Implementierung von deklarativen Regeln im CEP-System entfallen. Dies resultiert in einer Verbesserung der Reaktivität in Produktionsstätten und spart *Zeit* und *Kosten* ein.

Proaktive Maßnahmen. Zukünftige Events werden im CEP standardmäßig nicht berücksichtigt [42]. Eine proaktive und frühe Reaktion auf Events erhöht jedoch den Geschäftswert des Systems [32]. Mit der Erweiterung um ML lässt sich dies auch im CEP-Umfeld erreichen. Damit werden Vorhersagen und proaktives Handeln auf Basis historischer Eventdaten durch Schwellenbereiche möglich.

[42] setzen auf ein Conditional Density Estimation trainiertes ML-CEP-System, welches Eventdaten von Kamerasystemen und Sensoren erhält. Der Produktionsdurchlauf wird in Echtzeit auf anormale Veränderungen geprüft und mit Eintrittswahrscheinlichkeiten abgeglichen. Dieselbe Aufgabe findet sich als zentrales Thema der ACM DEBS 2017 Grand Challenge [53]. Zur Datenvorverarbeitung galt es Events auf Basis diskreter Zustände (z. B. Temperatur oder Druck der Produktionsmaschine) durch einen Clustering-Algorithmus zu gruppieren. Daraufhin musste ein Markov-Modell trainiert werden, um Eintrittswahrscheinlichkeiten von Anomalien in Echtzeit

zu berechnen. In allen Lösungsansätzen wurde mit Zustandsübergängen zwischen den Clustern gearbeitet. Das System benachrichtigt den Mitarbeiter sobald die Wahrscheinlichkeit über einem definierten Schwellenwert liegt.

[34] zeigen mit Hilfe eines Shapelet-basierten Ansatzes, dass sich die Vorhersagefähigkeit mit dem automatisieren Lernen von Regeln kombinieren lässt. Ein alternativer Ansatz [36] zielt auf die einheitliche Verwendung eines proaktiven Architekturstandards im industriellen Umfeld ab. Mit Hilfe eines Bayesschen Netzes und einer verbundenen Klassifizierung anhand von Entscheidungsbäumen werden in diesem Beispiel Probleme Zeit-terminierter Wartungsvorgänge unabhängig vom tatsächlichen Maschinenzustand angegangen. Durch eine CEP-Kombination mit einer ML-gestützten Echtzeitanalyse wird der Austausch bei faktischer Notwendigkeit eines Ersatzteils in Abhängigkeit der verbleibenden Lebensdauer angezeigt. Dies erlaubt eine entsprechende Optimierung des Wartungsintervalls.

Weitere Einsatzszenarien finden sich einerseits für intra-logistische Materialbewegungen wie der gezielten Überwachung von Förderbändern [63] sowie andererseits für extra-logistische Materialbewegungen wie z. B. im Straßenverkehr [8, 64, 65]. Dabei werden Eventdaten kontinuierlich über Sensoren, Kamerasystemen oder Induktionsschleifen erfasst, übertragen und zur Vorhersage u. a. von Verkehrsstaus herangezogen. [8] setzen dazu auf eine Adaptive Moving Window Regression, um ein kontinuierliches Lernen des CEP-Systems zu sichern. [64]. [65] nutzen hingegen Bayessche Netze, um probabilistische Complex Events vorherzusagen.

Die Proaktivität der aufgezeigten Ansätze führt vorrangig zur Optimierung der *Auslastung* der Fertigungs-, Wartungs- und Warenflussprozesse. Ebenso resultieren daraus *Kosteneinsparungen* durch proaktiv reduzierte Wartungsprozesse.

6 Diskussion und Fazit

Diskussion. Die unternommene Untersuchung zeigt, dass ein Einsatz von ML großes Potenzial für CEP besitzt. [66] bestätigen dies und bezeichnen den Einsatz von ML als bedeutenden Fortschritt in der Echtzeitauswertung von Events. Trotz der Bedeutung für die Forschungsdomäne fehlte bislang eine umfassende Betrachtung der unterschiedlichen Kombinationsansätze.

Auf Basis der durchgeführten Literaturrecherche konnte festgestellt werden, dass bei der Erweiterung von CEP um ML primär Supervised-Learning-Ansätze eingesetzt werden. Über die letzten Jahre finden sich aber vermehrt Kombinationen aus Supervised- und Unsupervised-Learning-Ansätzen. Dies birgt großes Potenzial für die Umsetzung proaktiven Verhaltens. Reinforcement Learning und weitere Alternativen sind hingegen kaum untersucht worden. Zukünftige Forschung könnte hier ansetzen, um bspw. die Wirksamkeit von belohnungsorientierten Ansätzen für das CEP zu untersuchen. Reinforcement Learning wird bereits in Multi-Agenten-Systemen verwendet. Da Event Processing Agents konzeptionell als autonome Agenten eines umfassenderen Event Processing Networks betrachten werden, liegt es nahe, die Wirksamkeit dieses Ansatzes auch in diesem Umfeld zu untersuchen.

In einer zweiten Synthese wurden mögliche Einsatzpotenziale von ML in CEP für Smart Factories betrachtet. Für die betriebswirtschaftlichen Zielbereiche von Kosten, Zeit, Flexibilität und Qualität ergaben sich sowohl (1) reaktive als auch (2) proaktive Verbesserungsansätze. Das Potenzial wird überwiegend in der Reduzierung der Zeit (d. h. in Bezug auf bessere Maschinenauslastung) und der Kosten (d. h. in Bezug auf Automatisierungsansätze) gesehen und adressiert. Bisherige Forschung zielte primär auf reaktive Maßnahmen ab, wobei gerade in den letzten beiden Jahren verstärkt auch proaktive Maßnahmen erforscht wurde.

Die Literaturanalyse zeigt, dass für reaktive Maßnahmen in Smart Factories vornehmlich auf Supervised Learning zurückgegriffen wird. Es handelt sich dabei insbesondere auf SVMs und künstlich neuronale Netze aus dem Bereich der Klassifizierung. Ansätze aus dem Unsupervised Learning sind bislang unterrepräsentiert, wobei jedoch erste Testversuche erfolgt sind. Combined Learning wiederum scheint für reaktive Maßnahmen hinsichtlich der fehlenden Verwendung vergleichsweise unbrauchbar. Dies steht im Gegensatz zu proaktiven Maßnahmen. Hier wird zunehmend mit Ansätzen aus dem Combined Learning gearbeitet. Primär erscheint eine Verbindung aus k-Means und dem Markov Model erfolgsversprechend. Auch Anwendungen aus dem Supervised Learning finden sich. Dabei werden überwiegend Bayessche Netze aus dem Bereich der probabilistischen Modelle eingesetzt. Ansätze aus dem Unsupervised Learning bestehen hingegen keine, was deren fehlende Eignung nahelegt.

Fazit und Ausblick. Aktuell werden einige Anwendungen für die kontinuierliche Echtzeitauswertung von Events auf CEP-Basis erprobt. Gerade im Umfeld von Smart Factories zeigen sich erste Umsetzungen primär in reaktiven Bereichen, wie der informationsgestützten Reparatur im Fehlerfall. Auch in Bezug auf Big Data werden Kombinationen aus CEP und ML als erfolgsversprechend angesehen [67]. Der Trend geht hierbei zu probabilistischen bzw. ungewissen Eventmodellen. Im Gegensatz zu deterministischen Eventmodellen sind dabei nicht alle Attribute bekannt oder mit Ungenauigkeiten versehen [67]. Erste Lösungsansätze fokussieren Bayessche und Markov-Netze allerdings mit bisher mäßigem Erfolg [68].

Zukünftige Forschung könnte hier ansetzen und über neue Ansätze wie bspw. das Online Machine Learning [69] weitere Verbesserungen erreichen. Beim Online ML steht die automatische Generierung der vormals manuell spezifizierten deklarativen CEP-Abfragen im Mittelpunkt, um sowohl eine verbesserte Eventerkennung und -verarbeitung zu erreichen als auch um Kosten zu sparen. Auch die Kombination von CEP mit Ansätzen aus dem Reinforcement Learning ist wie oben beschrieben vielversprechend. Gerade für Smart Factories kann im Hinblick auf die Optimierung der Maschinenlauffähigkeit und für das damit verbundenen Wartungsmanagement eine Verbesserung erfolgen. Dies lässt sich insbesondere durch präskriptive Analyse prognostizieren, wobei hier nicht die Maschine als Einzelkomponente betrachtet werden sollte, sondern die jeweils beste betriebswirtschaftliche Entscheidung im Gesamtunternehmenskontext [70].

Förderhinweis

Dieses Forschungs- und Entwicklungsprojekt wird mit Mitteln des Bayerischen Staatsministeriums für Wirtschaft, Landesentwicklung und Energie (StMWi) innerhalb des Förderprogramms "Informations- und Kommunikationstechnik" (Kennzeichen IUK536/003) gefördert und vom Projektträger VDI+VDE Innovation + Technik GmbH betreut.

Literaturverzeichnis

1. Yin, S., Kaynak, O.: Big Data for Modern Industry: Challenges and Trends. Proceedings of the IEEE 103, 143-146 (2015)
2. Bruns, R., Dunkel, J.: Complex Event Processing: Komplexe Analyse von massiven Datenströmen mit CEP. Springer Vieweg, Berlin (2015)
3. Luckham, D.: Event Processing for Business: Organizing the Real Time Enterprise. John Wiley & Sons, Inc., Hoboken, NJ (2011)
4. Kagermann, H., Wahlster, W., Helbig, J.: Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group. acatech (2013)
5. Akbar, A., Carrez, F., Moessner, K., Sancho, J., Rico, J.: Context-aware stream processing for distributed IoT applications. IEEE 2nd World Forum on Internet of Things, pp. 663–668. IEEE Press, New York (2015)
6. Turchin, Y., Gal, A., Wasserkrug, S.: Tuning complex event processing rules using the prediction-correction paradigm. Proceedings of the Third ACM International Conference on DEBS. ACM, New York (2009)
7. Mutschler, C., Philippsen, M.: Learning event detection rules with noise hidden Markov models. Proceedings of the 2012 NASA/ESA Conference on Adaptive Hardware and Systems, pp. 159–166. IEEE Press, New York (2012)
8. Akbar, A., Carrez, F., Moessner, K., Zoha, A.: Predicting complex events for pro-active IoT applications. IEEE 2nd World Forum on Internet of Things, pp. 327–332. IEEE Press, New York (2015)
9. Artikis, A., Paliouras, G., Portet, F., Skarlatidis, A.: Logic-based representation, reasoning and machine learning for event recognition. Proceedings of the Fourth ACM International Conference on DEBS, pp. 282–293. ACM, New York (2010)
10. Mehdiyev, N., Krumeich, J., Werth, D., Loos, P.: Sensor event mining with hybrid ensemble learning and evolutionary feature subset selection model. 2015 IEEE International Conference on Big Data, pp. 2159–2168. IEEE Press, New York (2015)
11. Mousheimish, R., Taher, Y., Zeitouni, K.: autoCEP: Automatic Learning of Predictive Rules for Complex Event Processing. In: Sheng, Q., Stroulia, E., Tata, S., Bhiri, S. (eds.) Service-Oriented Computing. ICSOC 2016. Lecture Notes in Computer Science, vol 9936, pp. 586–593. Springer, Cham, CHE (2016)
12. Widder, A., Ammon, R., Hagemann, G., Schönfeld, D.: An Approach for Automatic Fraud Detection in the Insurance Domain. 2009 AAAI Spring Symposium, pp. 98–100 (2009)
13. Samuel, A.L.: Some studies in machine learning using the game of checkers. IBM Journal of Research and Development 3, 210–229 (1959)
14. Bauernhansl, T.: Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma. In: Vogel-Heuser, B., Bauernhansl, T., Hompel, M.t. (eds.) Industrie

4.0 in Produktion, Automatisierung und Logistik, pp. 5-35. Springer Vieweg, Wiesbaden (2014)
15. Etzion, O., Niblett, P.: Event Processing in Action. Manning, Stamford, CT (2011)
16. Marsland, S.: Machine Learning: An Algorithmic Perspective. Chapman & Hall/CRC (2015)
17. Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, New York, NY (2014)
18. Wang, S., Chaovalitwongse, W., Babuska, R.: Machine Learning Algorithms in Bipedal Robot Control. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42, 728–743 (2012)
19. Ghahramani, Z.: Unsupervised Learning. In: Hutchison, D., Kanade, T., Kittler, J. (eds.) Advanced Lectures on Machine Learning, pp. 72–112. Springer, Berlin/Heidelberg (2004)
20. Kubat, M.: An Introduction to Machine Learning. Springer International Publishing, Cham, CHE (2017)
21. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (2017)
22. Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge, MA (2004)
23. Tokic, M.: Reinforcement Learning: Psychologische und neurobiologische Aspekte. Künstliche Intelligenz 27, 213–219 (2013)
24. vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on importance of rigour in documenting the literature search process. ECIS 2019, vol. 9, pp. 2206-2217, Verona (2009)
25. Cooper, H.M.: Organizing knowledge syntheses. A taxonomy of literature reviews. Knowledge in Society 1, 104–126 (1988)
26. Bhandari, S.: An adaptive event stream processing environment. Proceedings of the on SIGMOD/PODS 2012 PhD Symposium, pp. 57–62. ACM, New York (2012)
27. Margara, A., Cugola, G., Tamburrelli, G.: Towards automated rule learning for complex event processing. (2013)
28. Widder, A., Ammon, R., Schaeffler, P., Wolff, C.: Identification of suspicious, unknown event patterns in an event cloud. Proceedings of the 2007 Inaugural International Conference on DEBS, pp. 164–170. ACM, New York (2007)
29. Shepherd, C., Günter, H.: Measuring supply chain performance: current research and future directions. Behavioral Operations in Planning and Scheduling, pp. 105-121. Springer, Berlin, Heidelberg (2010)
30. Neely, A., Gregory, M., Platts, K.: Performance measurement system design: A literature review and research agenda. International Journal of Operations & Production Management 15, 80-116 (1995)
31. Gad, R., Kappes, M., Medina-Bulo, I.: Analysis of the Feasibility to Combine CEP and EDA with Machine Learning using the Example of Network Analysis and Surveillance. X Jornadas de Ciencia e Ingeniería de Servicios, (2014)
32. Schwegmann, B., Matzner, M., Janiesch, C.: preCEP: Facilitating Predictive Event-Driven Process Analytics. In: vom Brocke, J., Hekkala, R., Ram, S., Rossi, M. (eds.) Design Science at the Intersection of Physical and Virtual Design. DESRIST 2013. Lecture Notes in Computer Science, vol 7939, pp. 448–455. Springer, Berlin/Heidelberg (2013)
33. Patri, O.P., Panangadan, A., Sorathia, V., Prasanna, V.: Sensors to Events: Semantic Modeling and Recognition of Events from Data Streams. International Journal of Semantic Computing 10, 461–471 (2016)

34. Mousheimish, R., Taher, Y., Zeitouni, K.: Automatic Learning of Predictive CEP Rules: Bridging the Gap between Data Mining and Complex Event Processing. Proceedings of the 11th ACM International Conference on DEBS, pp. 158–169. ACM, New York (2017)
35. Metz, D., Karadgi, S., Müller, U., Grauer, M.: Self-Learning Monitoring and Control of Manufacturing Processes Based on Rule Induction and Event Processing. The Fourth International Conference on Information, Process, and Knowledge Management, pp. 88–92. IARIA (2012)
36. Engel, Y., Etzion, O.: Towards proactive event-driven computing. Proceedings of the 11th ACM International Conference on DEBS, pp. 125–136. ACM, New York (2017)
37. Borkowski, M., Fdhila, W., Nardelli, M., Rinderle-Ma, S., Schulte, S.: Event-based failure prediction in distributed business processes. Information Systems (2017)
38. Shahad, R.A., Bein, L.G., Saad, M. H., Hussain, A.: Complex event detection in an intelligent surveillance system using CAISER platform. International Conference on Advances in Electrical, Electronic and Systems Engineering, pp. 129–133. IEEE Press, New York (2016)
39. Baldoni, R., Montanari, L., Rizzut, M.: On-line failure prediction in safety-critical systems. Future Generation Computer Systems 45, 123–132 (2015)
40. Mehdiyev, N., Krumeich, J., Werth, D., Loos, P.: Determination of Event Patterns for Complex Event Processing Using Fuzzy Unordered Rule Induction Algorithm with Multi-objective Evolutionary Feature Subset Selection. 49th Hawaii International Conference on System Sciences, pp. 1719–1728. IEEE Press, New York (2016)
41. Skarlatidis, A., Paliouras, G., Artikis, A., Vouros, G.A.: Probabilistic Event Calculus for Event Recognition. ACM Transactions on Computational Logic 16, 11-47 (2015)
42. Christ, M., Krumeich, J., Kempa-Liehr, A.W.: Integrating Predictive Analytics into Complex Event Processing by Using Conditional Density Estimations. 20th International Enterprise Distributed Object Computing Workshop, pp. 287–294. IEEE Press, New York (2016)
43. Katzouris, N., Artikis, A., Paliouras, G.: Incremental learning of event definitions with Inductive Logic Programming. Machine Learning 100, 555–585 (2015)
44. Lang, J., Capik, Z.: Prediction based on hybrid method in complex event processing. 12th International Symposium on Applied Machine Intelligence and Informatics, pp. 315–320. IEEE Press, New York (2014)
45. Akbar, A., Khan, A., Carrez, F., Moessner, K.: Predictive Analytics for Complex IoT Data Streams. IEEE Internet of Things Journal 4, 1571–1582 (2017)
46. Roledene, S., Ariyathilaka, L., Liyanage, N., Lakmal, P., Bamunusinghe, J.: GeniBux - event based intelligent Forex trading strategy enhancer. IEEE International Conference on Information and Automation for Sustainability, pp. 1–6. IEEE Press, New York (2016)
47. Lee, O.-J., Jung, J.E.: Sequence Clustering-based Automated Rule Generation for Adaptive Complex Event Processing. Future Generation Computer Systems 66, 100–109 (2017)
48. Cuzzocrea, A., Folino, F., Guarascio, M., Pontieri, L.: Predictive Monitoring of Temporally-Aggregated Performance Indicators of Business Processes against Low-Level Streaming Events. Information Systems In Press, Accepted Manuscript, (2018)
49. Junior, M.R., Endler, M., Silva e Silva, F.J.D.: An on-line algorithm for cluster detection of mobile nodes through complex event processing. Information Systems 64, 303–320 (2017)
50. Sugiura, K., Ishikawa, Y., Sasaki, Y.: Grouping Methods for Pattern Matching over Probabilistic Data Streams. IEICE Transactions on Information and Systems E100D, 718–729 (2017)
51. Gander, M., Felderer, M., Katt, B., Tolbaru, A., Breu, R., A., M.: Anomaly Detection in the Cloud: Detecting Security Incidents via Machine Learning. In: Moschitti, A., Plank, B.

(eds.) Trustworthy Eternal Systems via Evolving Software, Data and Knowledge. EternalS 2012. Communications in Computer and Information Science, vol 379, pp. 103–116. Springer, Berlin/Heidelberg (2013)
52. Teixeira, P., Clemente, R.G., Kaiser, R.A., Viera Jr., D.A.: HOLMES: an event-driven solution to monitor data centers through continuous queries and machine learning. Proceedings of the Fourth ACM International Conference on DEBS, pp. 216–221. ACM, New York (2010)
53. Rivetti, N., Busnel, Y., Gal, A.: FlinkMan: Anomaly Detection in Manufacturing Equipment with Apache Flink: Grand Challenge. Proceedings of the 11th ACM International Conference on DEBS, pp. 274–279. ACM, New York (2017)
54. Chandrathilake, H.M.C., Hewawitharana, H.T.S., Jayawardana, R.S., Viduranga, A.D.D., Dilum Bandara, H.M.N., Marru, S., Perera, S.: Reducing computational time of closed-loop weather monitoring: A Complex Event Processing and Machine Learning based approach. Moratuwa Engineering Research Conference, pp. 78–83. IEEE Press, New York (2016)
55. Sejdovic, S., Hegenbarth, Y., Ristow, G.H., Schmidt, R.: Industry Paper: Proactive Disruption Management System: How not to be Surprised by Upcoming Situations. Proceedings of the 10th ACM International Conference on DEBS, pp. 281–288. ACM, New York (2016)
56. Preuveneers, D., Berbers, Y., Joosen, W.: SAMURAI: A batch and streaming context architecture for large-scale intelligent applications and environments. Journal of Ambient Intelligence and Smart Environments 8, 63–78 (2016)
57. Terroso-Sáenz, F., Cuenca-Jara, J., Gonzalez-Vidal, A.: Human Mobility Prediction Based on Social Media with Complex Event Processing. International Journal of Distributed Sensor Networks 12, 1–25 (2016)
58. Zhang, W., Hirzel, M., Grove, D.: AQuA: adaptive quality analytics. Proceedings of the 10th ACM International Conference on DEBS, pp. 169–180. ACM, New York (2016)
59. Peterson, E., To, M.A., Maag, S.: A novel online CEP learning engine for MANET IDS. 9th IEEE Latin-American Conference on Communications. IEEE Press, New York (2017)
60. Gungor, V.C., Hancke, G.P.: Industrial Wireless Sensor Networks: Challenges, Design Principles, and Technical Approaches. IEEE Transactions on Industrial Electronics 56, 4258-4265 (2009)
61. Neto, P.P., N. J.; Moreira, A. P: Accelerometer-based control of an industrial robotic arm. The 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 1192-1197 IEEE Press, New York (2009)
62. Mitchell, R., Chen, L.-R.: A survey of intrusion detection techniques for cyber-physical systems. ACM Computing Surveys 46, Article No. 55 (2014)
63. Beamon, B.M.: System reliability and congestion in a material handling system. Computers & Industrial Engineering 36, 673-684 (1999)
64. Akbar, A., Kousiouris, G., Pervalz, H., Sancho, J., Ta-Shima, P., Carrez, F., Moessner, K.: Real-Time Probabilistic Data Fusion for Large-Scale IoT Applications. IEEE Access 6, 10015–10027 (2018)
65. Wang, Y., Gao, H., Chen, G.: Predictive complex event processing based on evolving Bayesian networks. Pattern Recognition Letters In Press (2017)
66. Dayarathna, M., Perera, S.: Recent Advancements in Event Processing. ACM Computing Surveys 51, Article No. 33 (2018)
67. Flouris, I., Giatrakos, N., Deligiannikis, A., Garfalakis, M., Kamp, M., Mock, M.: Issues in complex event processing: Status and prospects in the Big Data era. Journal of Systems and Software 127, 217–336 (2017)

68. Alevizos, E., Skarlatidis, A., Artikis, A., Paliouras, G.: Probabilistic Complex Event Recognition: A Survey. ACM Computing Surveys 50, 71-101 (2017)
69. Peterson, E., To, M.A., Maag, S.: An online learning based approach for CEP rule generation. 8th IEEE Latin-American Conference on Communications, pp. 1–6. IEEE Press, New York (2016)
70. Wang, G., Gunasekaran, A., Ngai, E.W., Papadopoulos, T.: Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics 176, 98-110 (2016)

Sensor retrofit for a coffee machine as condition monitoring and predictive maintenance use case

Peter Burggräf[1], Johannes Wagner[1], Benjamin Koke[1], Kailashnath Manoharan[1]

[1] University of Siegen, Mechanical Engineering, Siegen, Germany
{peter.burggräf,johannes.wagner,benjamin.koke}@uni-siegen.de

Abstract. The concept of Industry 4.0 provides promising approaches to reduce downtime and increase overall equipment efficiency in manufacturing processes through interconnected devices in the industrial internet of things (IIoT). As the procurement of new IIoT-ready machines is costly, the retrofit of old machines can be an idea worth exploring. In this paper, we designed a simple experiment setup using affordable sensors and a coffee machine (due to the absence of machinery) to measure grinding vibrations and to predict the last coffee before grinder no-load. Microsoft Azure Machine Learning Studio was used to deploy machine learning techniques in order train prediction models. While prediction accuracy in this experiment was non-satisfactory, our results nonetheless indicate that retrofit is indeed a proper approach to make an older machine park smart, provided that sensors (especially their sample rate) are suitable for the application.

Keywords: Condition Monitoring, Machine Learning, Retrofit, Industry 4.0, Internet of Things

1 Introduction

With the rise of the concept of Industry 4.0 and the advent of data analytics and machine learning, **condition monitoring** to understand machine status and detect faults is considered as an important research topic [1]. In Industry 4.0, manufacturing systems include numerous sensors and interconnected devices (so-called cyber-physical systems) that form the industrial internet of things (IIoT) and facilitate the collection of large amounts of data. Data Analytics helps manufacturing firms to get actionable insights resulting in smarter decisions and better business outcomes [2]. The implementation of Machine Learning (ML) and Big Data may drive the next wave of innovation and may soon prove to be an unavoidable tactical move in achieving higher levels of optimization [3]. Some authors even argue that ML will play a major role in managerial decision-making [4].
Enabling communication and data collection in established manufacturing shop-floors, however, requires extensive capital investment in new machinery and IT infrastructure, which may prove to be a burden for small and medium-sized enterprises (SMEs). One alternative is the retrofit of older machines with sensor kits to collect machine data to

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

gain valuable insights. The purpose of this paper is to demonstrate the ease of retrofit initiatives by presenting an experiment setup using an affordable sensor kit and a coffee machine. The experiment involves detection of shifts in vibration frequency when the bean container runs empty to predict the last coffee. The remainder of this paper is structured as follows: first, we present relevant literature background for ML. Second, we present the experiment and results. Third, we present a discussion of the quality of results. Lastly, a conclusion and future outlook is presented.

2 Literature

ML analyses past data and makes predictions as well as decisions based on it. The machine learns to create its own logics and solutions with or without any form of human intervention [3]. There are different types of ML techniques relevant for real time scenario in a condition monitoring system. Supervised learning approaches use defined input-output tuples and require extensive training in order to make predictions for the future while unsupervised learning acts on information without pre-labeled output [5]. Classification and regression problems are the most frequently used methods in supervised learning whereas k-means clustering, auto-encoders and anomaly detection are popular in unsupervised learning [6].

Regression algorithms that are most commonly used in the domain of vibration monitoring studies include: Linear Regression [7] identifies a relation between dependent and independent variables. Bayesian Linear Regression [8] uses probability distributions rather than point estimates on generally small datasets. Neural Network Regression [9] approximate nonlinear functions of their inputs by adapting weights. Boosted Decision Tree Regression [10] creates a prediction model in the form of an ensemble of weak prediction models. With Decision Forest Regression [11], labeled set of inputs learn a general mapping which associates previously unseen independent test data with their correct continuous prediction.

As the experiment is designed as regression problem, the experiment setup (see next section) will focus on the use of regression algorithms.

3 Experiment

3.1 Setup

This experiment serves to approximate manufacturing condition monitoring by equipping a coffee machine with sensors. It is assumed that the same process applies to regular industrial applications in smart factories. Figure 1 shows the steps involved in the IoT Condition Monitoring use case of the coffee machine.

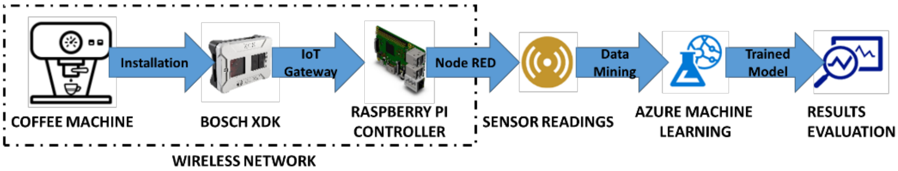


Figure 1. Coffee Machine IoT Use Case

A Bosch XDK device was mounted on the casing atop the bean grinder. Various sensor measurements (x, y and z acceleration, temperature, light, humidity, air pressure and orientation) were taken and sent to a Raspberry Pi via Wi-Fi. To reduce the amount of null measurements (machine idle), a threshold was set and measurements performed for 60 seconds. Using Node-RED, the data are transferred from the Raspberry Pi to a local server as comma-separated value (CSV) files. The CSV files were manually uploaded to **Microsoft Azure Machine Learning Studio** where data analytics and ML took place.

The raw dataset consisted of 49443 instances and 11 unique features (8 sensor measurements and 3 time stamps). The data was split into 70 percent for the training set and 30 percent for the testing. Since the dataset had over 95 percent of samples in abundant class (Idle Condition of Coffee Machine) and less than 5 percent being the rare class (Coffee Intake), it was essential to resample the training and testing datasets. Synthetic Minority Oversampling technique was used to increase the size of the rare class. Another major challenge in data preprocessing is to resolve dataset imbalance. Joining data is another complex process in the initial data preparation due to large number of missing values. Those were cleaned using Replace With Median and Probabilistic PCA (principal component analysis) approaches. Boosted Decision Tree Regression, Two-class Neural Network, Decision Forest and Bayesian Linear Regression models were trained and scored in this experiment using the feature **magnitude of vibration** as scored labels. For model testing, we deployed the model as Excel web service and provided time stamps for a typical intake cycle as input variables.

3.2 Results

With an accuracy of 62.06% (as measured by the coefficient of determination), Boosted Decision Tree Regression outperformed the other algorithms for this particular dataset. The performance of the learning algorithm strongly relies on the nature of the dataset. Other algorithms have also performed well and could be feasible options as they display accuracy slightly above 50%. When using these algorithms in regression models, slight changes in conditions between different datasets lead to wide temporal variations in the weights associated with individual features. So different temporal datasets lead to the vastly different relative importance of features, which makes the analysis of trends difficult.

4 Discussion

Good accuracy for trained models depend on several factors such as the algorithm selection, feature selection, quality, sensibility and size of data. In a classifier, 50% accuracy is considered as random guessing whereas in regression, 50% accuracy can be considered good or bad depending on the data. For the given experiment, the dataset consisted of 2472 useful data points (5%, see section 3.1). A regular coffee intake cycle held around 600 data points (average). Thus, predicting a cycle with 62% accuracy results in 228 inaccurate data points. To detect minor shifts in vibration frequency, the accuracy of 62.06% must therefore be considered non-satisfactory.

An explanation for this deficit can be found upon closer examination of the gathered data. The time stamps reveal that there is an average of 1 and a maximum of 2 measurements per second which seems to be insufficient to properly represent higher-frequency vibrations.

5 Conclusion

In this paper, we described an experiment setup to demonstrate the ease of machine retrofit for Industry 4.0 compatibility. We attempted to predict the last coffee before coffee bean depletion by collecting vibration data from a coffee machine using a Bosch XDK sensor kit and deploying ML algorithms on Microsoft Azure.

The experiment results indicate that while algorithm selection for this experiment can be considered satisfactory, data collected from this sensor are not accurate enough for this application. Using a different type of accelerometer (e.g. piezo resistive) in combination with a microcontroller with sufficiently high sample rate might help in getting better results for this use case. Therefore, as a next step, we will re-attempt the experiment using improved data acquisition. Nonetheless, we believe that this experiment highlighted that there is efficiency potential for manufacturing in retrofitting old machines through the advancement of ML and IIoT.

References

1. Moraru, A., Pesko, M., Porcius, M., Fortuna, C., Mladenic, D.: Using machine learning on sensor data. CIT 18, 341–347 (2010)
2. Lee, J., Bagheri, B., Kao, H.-A.: Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. In: Proceeding of Int. Conference on Industrial Informatics (INDIN) 2014, pp. 1–6 (2014)
3. Hoogendoorn, M., Funk, B.: Reinforcement Learning to Provide Feedback and Support. In: Machine Learning for the Quantified Self, pp. 203–214. Springer (2018)
4. Burggräf, P., Wagner, J., Koke, B.: Artificial intelligence in production management: A review of the current state of affairs and research trends in academia. In: 2018 International Conference on Information Management and Processing (ICIMP), pp. 82–88. IEEE (2018)
5. Alpaydin, E.: Introduction to machine learning. MIT Press, Cambridge, MA, London (2004)
6. Kubat, M.: An introduction to machine learning. Springer, Cham (2015)

7. Aherwar, A., Unune, D., Pathri, B., kishan, J.: Statistical and Regression Analysis of Vibration of Carbon Steel Cutting Tool for Turning of EN24 Steel Using Design of Experiments. IJMECH 3, 137–151 (2014)
8. Ye, Z., Chen, N., Tsui, K.-L.: A Bayesian Approach to Condition Monitoring with Imperfect Inspections. Qual. Reliab. Eng. Int. 31, 513–522 (2015)
9. Schlechtingen, M., Ferreira Santos, I.: Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mech. Syst. Sig. Process. 25, 1849–1875 (2011)
10. Burges, C.J.C.: From ranknet to lambdarank to lambdamart: An overview. Learning 11, 81 (2010)
11. Criminisi, A.: Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. FNT in Computer Graphics and Vision 7, 81–227 (2011)

Stakeholder-Analyse zum Einsatz IIoT-basierter Frischeinformationen in der Lebensmittelindustrie

Stephanie Vonholdt[1], Gunnar Stevens[2], Darius Becker[1]

[1] Hochschule Bonn-Rhein-Sieg, Wirtschaftswissenschaften, Sankt Augustin, Deutschland
Stephanie.Vonholdt@h-brs.de, Darius.Becker@smail.wis.h-brs.de
[2] Universität Siegen, IT-Sicherheit und Verbraucherinformatik, Siegen, Deutschland
Gunnar.Stevens@uni-siegen.de

Abstract. Eine Herausforderung bei der Implementierung des industriellen Internet of Things (IIoT) besteht darin, Mehrwerte in Wertschöpfungsketten zu identifizieren, um darauf aufbauend Lösungen nutzerzentriert zu gestalten. Dieser Beitrag stellt das Forschungsprojekt FreshIndex vor, bei dem diese Herausforderung durch eine Kombination aus Stakeholder-Analyse und User-Centered-Design-Methoden adressiert wurde. Ziel des Projekts ist es, eine IIoT-basierte Lösung zum Monitoring der Kühlkette in der Lebensmittelindustrie zu entwickeln. Hierzu ist es wichtig zu wissen, welche Nutzer/-innen mit den Daten in Berührung kommen und welche Erfahrungen, Fähigkeiten, Anforderungen und Wünsche sie mitbringen. Die Berücksichtigung dieser Aspekte ist relevant für den Erfolg der Konzeption, Implementierung und des Betriebs eines IIoT-Systems. So können nützliche und handhabbare Produktideen generiert und Anwendungen gestaltet werden, die von Mitarbeiter/-innen und Konsument/-innen angenommen werden. IIoT schließt somit die lokale Verwendbarkeit von Daten entlang der Wertschöpfungskette ein und beschränkt sich nicht auf zentrale Verfügbarkeit von Daten.

Keywords: Stakeholder-Analyse, User-Centered Design, IIoT (Industrial Internet of Things), EPCIS (Electronic Product Code Information Services).

1 Einleitung

Bei der Betrachtung des Internet of Things (IoT), insbesondere beim industriellen Internet of Things (IIoT) wird häufig eine ausschließlich technologische Sichtweise verwendet, wie z. B. von Gubbi et al. [1–3]. Dabei sind es die Nutzer/-innen, die ihren Input in das IIoT-System einbringen, mit den Daten arbeiten und von den Ergebnissen profitieren können. Um IIoT-Systeme mitarbeiter- und kundenorientiert zu gestalten, muss der lokale, sozio-technische Kontext erhoben werden [4]. In dieser Studie zeigen wir exemplarisch an einem IIoT-Projekt zur Dokumentation der Kühlkette von Fleisch, wie eine Stakeholder-Analyse durchgeführt und damit nutzerzentrierte Lösungen entwickelt werden können. Das Projekt, in dem die Analyse durchgeführt wurde, adres-

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

siert die akute gesellschaftliche Herausforderung der Reduzierung von Lebensmittelabfällen. Jüngsten Studien zufolge werden in der EU jährlich Nahrungsmittel im Wert von 143 Milliarden Euro weggeworfen [5]. Ein wesentlicher Teil der Abfälle entsteht entlang der immer komplexer werdenden Lieferkette durch Nichterfüllung der Qualitäts- und Sicherheitsanforderungen, Verzögerungen im Prozess oder Ablauf des Verfallsdatums ohne Verkauf oder Verbrauch der Ware [6]. Im Jahr 2015 verabschiedete die UN-Generalversammlung Ziele für eine nachhaltige Entwicklung bis 2030, die eine Verpflichtung zur Halbierung der Nahrungsmittelverschwendung pro Kopf beinhalteten [7]. Ziel des Projektes ist die Entwicklung eines dynamischen Haltbarkeitskriteriums für Lebensmittel. Dazu werden die bestehenden EPCIS (Electronic Product Code Information Services) [8] um Informationen zu Lagerbedingungen und Produktqualität entlang der gesamten Wertschöpfungskette erweitert. In der Wertschöpfungskette von Lebensmitteln entstehen eine Vielzahl von Sensordaten. Die Temperaturüberwachung ist die wichtigste Anforderung beim Management einer Kühlkette. Das IIoT bietet neue Möglichkeiten, die Temperatur von Produkten digital zu verfolgen. Durch eine dynamische Anpassung des Haltbarkeitsdatums bei Übererfüllung der Kühlanforderungen, kann die vorzeitige Entsorgung von Lebensmitteln deutlich reduziert werden. Digitale Plattformen verbessern den Datenaustausch zwischen den involvierten Unternehmen und Endverbraucher/-innen und bringen damit zusätzliche Transparenz und Sicherheit. Um IIoT nutzstiftend verwenden zu können, muss jedoch der lokale Kontext der Informationserzeugung, -verarbeitung und -nutzung berücksichtigt werden.

2 Angewendete Methoden zur Analyse der Anforderungen

Im Rahmen der Forschungs- und Entwicklungsarbeit wird aktuell mit den verschiedenen, für den Projekterfolg relevanten Akteuren eine Anforderungsanalyse durchgeführt. Hierzu muss die, zunächst trivial erscheinende Frage geklärt werden, welche Stakeholder bei der Produktion, Distribution und Konsumption von den Daten der Kühlkette beteiligt sind. Da die verschiedenen Stakeholder unterschiedliche Ansprüche an die IIoT-Lösung haben und das Themenfeld noch neu ist, wurden im ersten Schritt qualitative Erhebungen durchgeführt. Dabei erfolgte die Auswahl der Interviewpartner/-innen auf Basis des Schneeballsystems, sowie einer ergänzenden, unternehmensübergreifenden Analyse des Wertschöpfungsprozesses.

2.1 Stakeholder-Analyse

Im Zuge der Stakeholder-Analyse wurde eine detaillierte Stakeholder-Übersicht ausgearbeitet, die einen Überblick darüber gibt, welche Personen potenzielle Nutzer/-innen sein könnten [Abbildung 1]. Der Begriff Stakeholder wurde gewählt, da es zunächst noch herauszufinden gilt, wer zukünftig Nutzer/-in der Frischedaten sein wird. Die Übersicht ist in englischer Sprache, da das Projekt im internationalen Kontext stattfindet. Diejenigen Stakeholder, die im Fokus des Projekts stehen und für dessen Umsetzung erforderlich sind, wurden im innersten Kreis platziert. Weitere Stakeholder, die im Rahmen des Projekts erforscht werden, aber für die Projektziele nicht erforderlich sind, befinden sich im mittleren Kreis. Die Stakeholder, die im Projekt nicht untersucht werden, befinden sich im äußeren Kreis. Dies können Stakeholder sein, die nicht direkt in den Prozess eingebunden sind, aber den Erfolg des Produktes in der Zukunft beeinflussen können. Es ist wichtig sie zu analysieren, aber erst zu einem späteren Zeitpunkt.

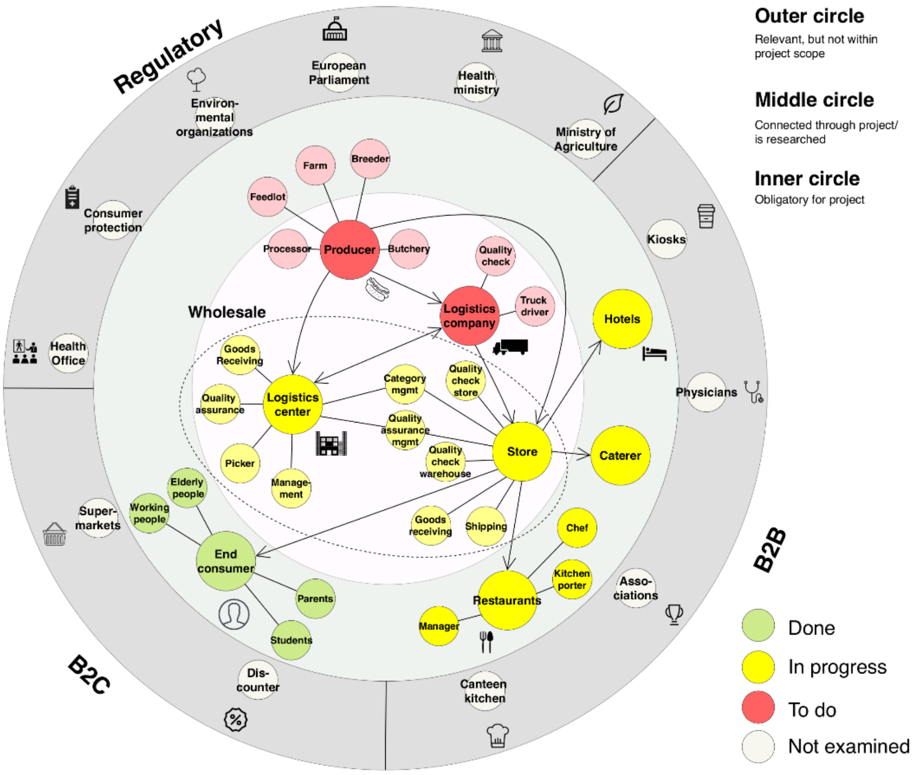


Abbildung 1: Stakeholder-Übersicht

Anschließend wurden die Stakeholder durch Pfeile in der Reihenfolge ihres Auftretens in der Wertschöpfungskette miteinander verbunden. Dies hilft, den Prozess lückenlos zu betrachten, einschließlich aller Schnittstellen und beteiligten Personen. Weitere Stakeholder wurden auch durch Empfehlungen der Interviewteilnehmer identifiziert. Die Stakeholder-Übersicht bietet eine Möglichkeit, mit Projektmitgliedern sowie mit Externen über die Relevanz bestimmter Stakeholder zu diskutieren. Sie hilft, einen Fokus zu setzen, aber auch Interessensgruppen, die nicht im Zentrum des Projekts stehen, in Zukunft nicht zu vergessen. Darüber hinaus ist sie ein lebendiges Instrument, das iterativ weiterwachsen kann und soll. Wenn ein neuer Stakeholder in den Erhebungen identifiziert wird, wird er in der Stakeholder-Übersicht platziert und muss so direkt im Projektkontext bewertet werden. Mit Ampelfarben, die signalisieren welche Bereiche untersucht wurden, hilft sie auch, den Überblick über To-Dos zu behalten.

2.2 Qualitative Forschung

Basierend auf der Stakeholder-Übersicht haben wir die Stakeholder mit Methoden aus dem User-Centered Design, wie Site Visit, Contextual Inquiry und semistrukturierte Interviews untersucht, um den jeweiligen Kontext der IIoT-Infrastruktur qualitativ zu erforschen [9]. Ziel der Forschung war es Personas und Kontextszenarien zu generieren, die Aufschluss über ihre Tätigkeit, Bildung, Erfahrung, IT-Nutzung, Ziele, Prozessintegration und Wünsche an dynamische Frischedaten geben. Dies hilft, die Bedürfnisse der möglichen zukünftigen Nutzer/-innen zu verstehen. Für jeden Stakeholder wurden geeignete Methoden zur Erhebung seiner Anforderungen gewählt. Im Bereich der Endverbraucher/-innen wurden 22 qualitative Interviews entlang semistrukturierter Leitfäden geführt. Endverbraucher/-innen kommen in sehr unterschiedlichem Kontext mit dem Thema Frische in Kontakt. Diese Unterschiede galt es hier zu erheben. Mitarbeiter/-innen hingegen befassen sich in einem ganz bestimmten Kontext mit Frische. Im Rahmen von Beobachtungen mit Fotodokumentation im Großhandel wurden Einblicke in die Arbeitsplatzgestaltung und -prozesse von Mitarbeiter/-innen gewährt, die Aufschluss über die Arbeitskontexte gaben. Anschließend wurden Interviews mit 11 Mitarbeiter/-innen geführt, die in unterschiedlichen Kontexten des Bereichs Frische arbeiten. Im B2B-Kundensegment HoReCa (Hotels, Restaurants und Caterer), der Hauptzielgruppe des Großhandels gab ebenfalls die Kombination von 6 Interviews mit Beobachtungen, erweiterte Einblicke in den Arbeitskontext. Produzenten und Logistikunternehmen ziehen den geringsten Nutzen aus Frischedaten der zurückliegenden Wertschöpfungskette und zeigten sich daher weniger interessiert, an der Forschung für dieses Produkt teilzunehmen. Für weitere Projekte bedeutet dies, dass diejenigen Stakeholder, die nicht so offensichtlich von der IIoT-Lösung profitieren, früh einbezogen werden müssen, da auch sie für den Erfolg relevant und schwieriger zu motivieren sind. Mittels qualitativer Inhaltsanalyse wurden die Erhebungen ausgewertet und Personas generiert. Ein Design-Thinking-Workshop mit weiteren Teilnehmern aus der Wertschöpfungskette findet am 25. September 2018 statt. Die Personas werden dort diskutiert und mit neuen Ideen ergänzt. So können verschiedene Sichtweisen zu den Stakeholdern erhoben und die ersten Ergebnisse verifiziert werden. Darüber hinaus werden dort Lösungsansätze in einem kreativen Umfeld erarbeitet. Im nächsten Schritt werden die Erkenntnisse über die potenziellen Nutzer/-innen zusammengefasst. Die Lösungsansätze werden hinsichtlich ihrer Potenziale und Realisierbarkeit bewertet und für die nutzerorientierte Prototypentwicklung verwendet.

Referenzen

1. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems. 29, 1645–1660 (2013).
2. Bandyopadhyay, D., Sen, J.: Internet of things: Applications and challenges in technology and standardization. Wireless Personal Communications. 58, 49–69 (2011).
3. Atzori, L., Iera, A., Morabito, G.: The internet of things: A survey. Computer networks. 54, 2787–2805 (2010).

4. Veryzer, R.W., Borja de Mozota, B.: The impact of user-oriented design on new product development: An examination of fundamental relationships. Journal of product innovation management. 22, 128–143 (2005).
5. Stenmarck, Å., Jensen, C., Quested, T., Moates, G., Buksti, M., Cseh, B., Juul, S., Parry, A., Politano, A., Redlingshofer, B.: Estimates of European food waste levels. IVL Swedish Environmental Research Institute (2016).
6. Kreyenschmidt, J., Albrecht, A., Braun, C., Herbert, U., Mack, M., Roissant, S., Ritter, G., Teitscheid, P., Ilg, Y.: Food Waste in der Fleisch verarbeitenden Kette: Um Lebensmittelverluste zu minimieren, sind Handlungen entlang der Kette Fleisch notwendig. Fleischwirtschaft. 93, 57–63 (2013).
7. Nationen, V.: Resolution der Generalversammlung, verabschiedet am 25. September 2015. Transformation unserer Welt: die Agenda. 2030, (2015).
8. GS1 Germany GmbH: EPCIS - GS1 Germany, https://www.gs1-germany.de/gs1-standards/datenaustausch/epcis/ (Accessed: 10.09.2018).
9. Vredenburg, K., Mao, J.-Y., Smith, P.W., Carey, T.: A survey of user-centered design practice. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 471–478. ACM (2002).

Towards a Framework for Predictive Maintenance Strategies in Mechanical Engineering – A Method-Oriented Literature Analysis

Nina Wiedemann[1], Friedemann Kammler[1], Andreas Varwig[1], and Oliver Thomas[1]

[1] University of Osnabrück, Osnabrück, Germany
{nwiedemann,fkammler,anvarwig,othomas}@uni-osnabrueck.de

Abstract. Industrial machines are amongst Germany's main export products and contribute to the increasing revenue of Mechanical Engineering. However, in the course of globalization, services for such machines have become costly and inflexible due to long distances between vendors and customers. Consequently, companies seek to avoid unexpected failures and long down times by the development of data-based "smart" service solutions, including Predictive Maintenance (PM). In contrast to reactive or preventive measures, PM refers to the proactive planning of required maintenance services based on data sampled from the machinery. Although PM has been conceptualized decades ago and various methods have been proposed ever since, there is no standard strategy. By analyzing existing literature, we shed light on the knowledge base in PM. We provide an overview of methods and discuss their respective context, including preconditions and applications. Our work constitutes a first step towards a framework that guides the implementation of PM-strategies.

Keywords: Predictive Maintenance, Service Strategies, Predictive Analytics, Literature Review, Mechanical Engineering

1 Motivation

Mechanical Engineering is a leading industry in Germany that generates most of its turnover on international markets. However, the traditional product business is under pressure of global competition, that also spreads to other revenue streams such as the spare part business [1]. In response, companies start to transform their business strategies, offering supplementary services or even "servitizing" aspects of the product business [2-3]. A key challenge to the gain in profitability of repair and maintenance activities are costs involved with the unexpected failure of machines in distant countries, e.g. high traveling expenses as well as opportunity costs resulting from long down times. Digital technologies offer improvements to this situation, since they allow for the remote identification of current and upcoming service needs.

One strategy is the analysis of sensor data from installed machines, in order to diagnose degradation and to predict the remaining useful lifetime. Such concepts contribute to the vision of "Predictive Maintenance" (PM), aiming to avoid both

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

unexpected failures of machines, as well as the opportunity losses of only partially worn spare parts [4-5]. In recent years, companies of various fields have started to implement PM systems, but often the proposed strategies are highly individual, focusing on a specific application scenario. The field of PM is thus very diverse, comprising a range of complex methods that is unstructured and difficult to grasp for non-experts. Therefore, we believe that research is required to facilitate the implementation of a PM strategy, in particular to unify and standardize existing approaches. In this contribution, we aim to provide a consolidated methodical overview based on an extensive literature study, in order to build a framework for PM methods in a next step.

2 Design and Results of our Study

We have conducted a structured literature study on articles addressing Predictive Maintenance in respect to the guidelines of Webster & Watson [6]. We firstly focused on articles in reputable information systems (IS) journals according to the A, B and C categories of the VHB-Jourqual3-ranking. However, PM articles are often highly methodical and are therefore also published in other domains. Thus, we complemented a general query via the paper hosts Google Scholar, Sciencedirect and Scopus, narrowing the extensive results to the 200 most relevant articles. During both reviews, we used the search term "Predictive Maintenance". By the analysis of title, abstract and keywords, we retrieved and categorized 87 relevant articles that describe concepts, methods or applications of predictive maintenance strategies.

Our literature study yields information on research trends in the field of PM. When the topic emerged around 1960, articles were highly conceptual, but over the years this focus increasingly shifted towards methodical and applicational papers. In particular, this became evident in the analysis of utilized data, where 16 of the 21 papers published before 1998 (76%) did not rely their contributions on any data. Apparently, PM research experienced a paradigm shift then, with only 8% articles after 1998 testing their strategies on neither real nor simulated data. In the last ten years, more than half of the publications deal with real data and thus with actual use cases, indicating that now PM has advanced from a conceptual challenge to a feasible business strategy. We believe that this development has been strongly encouraged by the advancement of methods. Fig. 1 presents the result of our methodological classification of the literature, which is an overview of identified method groups, their subclasses and specific methods. For a long time, only functional approaches that model the degradation process of a unit over time have been employed. From 1990 on, the methodical spectrum enlarged. As in other fields, we identified the increasing popularity of Machine Learning and Data Science for PM purposes. The interest in such methods can be linked to parallel advances in hardware that enabled the efficient use of methods such as Artificial Neural Networks (ANNs). By now, research employing reliability functions accounts for 32% of our literature base, whereas ANNs have reached 23%, followed by modern Time Series Analysis (15%).

Further, the analysis provides first insights on the context of application, benefits and drawbacks of each method. Firstly, the articles describe utilizable methods for

Figure 1. Overview of methods employed in Predictive Maintenance articles

particular tasks: For example, classification methods such as ANNs or Decision Trees are well suited for a simple status classification or failure recognition [7-8], [9]. If the residual life or future malfunctions should be predicted, a method that incorporates temporal information is required (e.g. Regression Analysis [10], Markov Models [11], [12] or Kalman Filtering [13-14]). Secondly, the method is highly dependent on the properties of the data. The analysis of event logs for instance can be conducted by text mining methods as in [15-17]. Similarly, image processing can now be performed efficiently by Convolutional Neural Networks (CNNs). [18] and [19] apply CNNs on thermographic images for failure recognition. But not only does the general data format influence the choice of method, other features such as scale or the proportion of outliers must be considered. Filtering methods might be necessary to smooth the data before further processing. For example, Yiwei et al. include Kalman filtering in their framework to track an aircraft's health state [20]. Another selection criterion can be the transparency of the inner mechanism of a method. For example, imagine that sensors are installed throughout a machine, and its failure is being predicted based on the data received. In order to execute maintenance tasks quickly, it would be helpful to isolate particular sensors that identify root causes of the problem. Consequently, methods providing high transparency during the information retrieval (e.g. Decision Trees) might be preferred in such scenarios, in contrast to ANNs which are often referred to as black boxes. Because of this, but also considering the problem of overfitting in complex function approximations such as ANNs, it is important to be aware of multiple methods and carefully compare "trending" with simple approaches.

3 Discussion

Our study exemplifies insights on methodology from the literature review, with more decision criteria yet to be analyzed, such as the industry or the relevance of runtime performance. Considering the plurality of influences and the diversity of use cases, it can already be concluded that this large variety of methods is inevitable. Nevertheless, it is crucial to shed light on the smorgasbord of methods. Our literature study does not cover all research extensively, but provides a structured overview of a multitude of methods and exemplary studies. We regard this as a suitable starting point to a framework for future PM implementations. However, the methodological point of view is not sufficient, but must be embedded in a holistic service and business model. The application scenario, including production, the exact execution of maintenance work and involved costs must be considered as well. PM necessitates high flexibility and cooperation between service providers and customers. Thus, we see further research need in the implementation of use cases and required information systems as well as the elaboration of new business strategies.

The ongoing research aims at the extensive classification of methods. We strive for the development of an information hub that allows for the requirements-based recommendation and choice of methods. This way, we hope the structured report on existing concepts and methods constitutes a first step to lower existing barriers for investing in PM, and enables more companies to seize the opportunity of innovative data-driven service strategies.

References

1. VDMA (2017) Maschinenbau in Zahl und Bild 2017. Frankfurt am Main, Verband deutscher Maschinen- und Anlagenbauer
2. Roland Berger (2014) Evolution of service: Capturing value pockets in the service business.
3. VDMA e.V.; McKinsey & Company (2014) Zukunftsperspektive deutscher Maschinenbau: Erfolgreich in einem dynamischen Umfeld agieren

4. Furtak, Simon; Avital, Michel; Ulslev Pedersen, Rasmus: Sensing the Future: Designing Predictive Analytics with Sensor Technologies. 23rd European Conference on Information Systems (2015)
5. Hashemian, H. M.; Bean, Wendell C. (2011) State-of-the-Art Predictive Maintenance Techniques*. IEEE Transactions on Instrumentation and Measurement, 60 (10):3480–3492.
6. Webster, Jane; Watson, Richard T. (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26 (2):xiii–xxiii.
7. Almeida, C. A. L., Braga, A. P., Nascimento, S., Paiva, V., Martins, H. J., Torres, R., Caminhas, W. M.: Intelligent thermographic diagnostic applied to surge arresters: a new approach. IEEE Transactions on Power Delivery, 24(2), 751-757 (2009)
8. Berecibar, M., Devriendt, F., Dubarry, M., Villarreal, I., Omar, N., Verbeke, W., Van Mierlo, J.: Online state of health estimation on NMC cells based on predictive analytics. Journal of Power Sources, 320, 239-250 (2016)
9. Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A.: Machine learning for predictive maintenance: A multiple classifier approach. IEEE Transactions on Industrial Informatics, 11(3), 812-820 (2015)
10. Onanena, R., Oukhellou, L., Candusso, D., Same, A., Hissel, D., Aknin, P.: Estimation of fuel cell operating time for predictive maintenance strategies. international journal of hydrogen energy, 35(15), 8022-8029 (2010)
11. Cartella, F., Lemeire, J., Dimiccoli, L., Sahli, H.: Hidden semi-Markov models for predictive maintenance. Mathematical Problems in Engineering (2015)
12. Lee, S., Li, L., Ni, J.: Online degradation assessment and adaptive fault detection using modified hidden Markov model. Journal of Manufacturing Science and Engineering, 132(2) (2010)
13. Abdennadher, K., Venet, P., Rojat, G., Rétif, J. M., Rosset, C.: A real-time predictive-maintenance system of aluminum electrolytic capacitors used in uninterrupted power supplies. IEEE Transactions on Industry Applications, 46(4), 1644-1652 (2010)
14. Yang, S. K.: An experiment of state estimation for predictive maintenance using Kalman filter on a DC motor. Reliability engineering & system safety, 75(1), 103-111 (2002)
15. Makanju, A. A., Zincir-Heywood, A. N., Milios, E. E.: Clustering event logs using iterative partitioning. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 1255-1264 (2009)
16. Xu, W., Huang, L., Fox, A., Patterson, D., Jordan, M. I.: Detecting large-scale system problems by mining console logs. In Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles, 117-132 (2009)
17. Sipos, R., Fradkin, D., Moerchen, F., Wang, Z.: Log-based predictive maintenance. In Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, 1867-1876 (2014)
18. Huda, A.N., Taib, S.: Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment. Applied Thermal Engineering, 61(2): 220–227 (2013)
19. Shafi'i, M.A., Hamzah, N.: Internal fault classification using artificial neural network. Power Engineering and Optimization Conference (PEOCO), 2010 4th International, 352–357 (2010)
20. Yiwei, W. A. N. G., Christian, G. O. G. U., Binaud, N., Christian, B. E. S., Haftka, R. T.: A cost driven predictive maintenance policy for structural airframe maintenance. Chinese Journal of Aeronautics, 30(3), 1242-1257 (2017)

Development of a matching platform for the requirement-oriented selection of cyber physical systems for SMEs

Tingni Xu[1], Anne Bernardy[2], Matthias Bertling[1], Peter Burggräf[1,3], Volker Stich[2], and Matthias Dannapfel[1]

[1] Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen, Germany
{T.Xu,M.Bertling,P.Burggraef,M.Dannapfel}@wzl.rwth-aachen.de
[2] FIR Institute for Industrial Management, RWTH Aachen University, Aachen, Germany
{Anne.Bernardy,Volker.Stich}@fir.rwth-aachen.de
[3] Chair for International Production Engineering and Management, University of Siegen, Siegen, Germany
{Peter.Burggraef}@uni-siegen.de

Abstract. This paper addresses the challenge of a systematic requirement-oriented configuration and selection of cyber physical systems (CPS) for SMEs. As the key technologies of realizing the digitalization and interconnection of production processes, manufacturing companies have realized the potential benefits brought by CPS. However, due to the complexity and fast development of CPS technology, it is difficult for SMEs, which lack expertise and financial resources, to select the appropriate CPS technologies meeting both functional and financial requirements. To overcome the issue, an online matching platform is developed to let SMEs express their needs and assist them conceptualize the individual CPS. This paper presents the matching methodology of the matching platform, which can not only match technical characteristics but also evaluate economic potentials. Then, it was demonstrated by a tracking and tracing use case in the end-of-line assembly of a small-sized German electric automobile manufacturer.

Keywords: cyber physical systems, matching, evaluation, SMEs

1 Introduction

Faced with the growing digitalization trend, manufacturing companies have realized the potential benefits of adopting cyber-physical systems (CPS) in their production processes. The function of CPS is characterized by the connection of physical and computational entities usually with feedback loops, which enables higher efficiency and flexibility in production monitoring and control [1-5]. However, due to lack of overview of all available CPS technology in the market and technical expertise, it lasts especially for SMEs a long time until CPS are conceptualized and implemented. In this

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

paper, a matching platform aimed at SMEs is developed and demonstrated, which supports the requirement-oriented selection of CPS.

2 State of the Art

In the literature, there are no specific methods but several general methods, which can be adopted to conceptualize CPS such as the guideline to design and construct technical systems and products (VDI 2221) [6]. However, it focuses a new development of technical systems, while implementation of CPS is usually a retrofit of existing infrastructure and a selection of available solutions. Thus, frame conditions and matching methodology need to be supplemented. Value benefit analysis and technical-economic evaluation method according to VDI 2225 [7–10] can contribute to functional or economic evaluation of CPS technologies. However, they fail in considering the two aspects in one framework. Therefore, a combined methodology to match technical features and evaluate economic potentials of CPS needs to be developed.

3 Development of CPS Matching Methodology

In this chapter, the developed CPS matching methodology including functional matching logic and economic evaluation model is presented. The methodology is then implemented in an online CPS matching platform: goo.gl/UmV7n8. The CPS matching platform will be introduced in more details through the case study in Chapter 4.

3.1 Functional Matching Logic

The approach to match user requirements and CPS technology functions is developed according to the VDI guideline 2221 [6] (see Figure 1). CPS technologies are broken down into six categories (actuators, sensors, transmission technologies, IT-Infrastructure, data processing, and HMI) to structure the components of CPS [11]. The functional fulfilment of available CPS technologies is assessed by a structural evaluation scheme containing all requirement specifications with a grading from 0 to 4 (see Figure 1) [12]. It enables user requirements and CPS technologies to be evaluated in a uniform scale, so that the matching of a CPS technology (or technology combinations) and requirements can be easily identified.

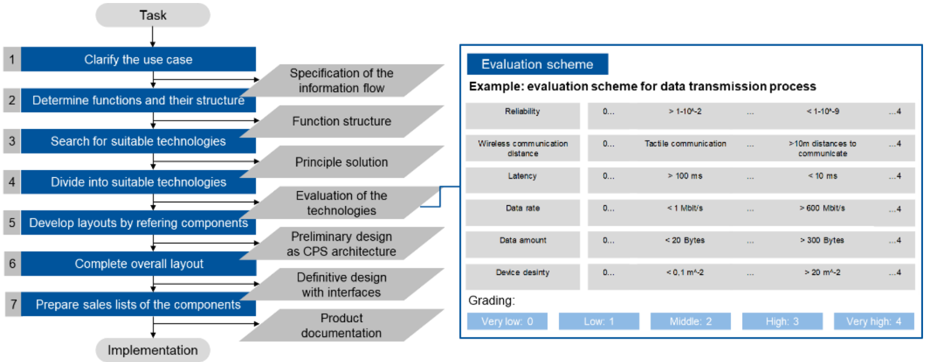


Figure 1. The overview of matching process and evaluation scheme for CPS functional fitting

3.2 Economic Evaluation Model

After a functional fitting, the matching platform can further assist user companies with economic evaluation of CPS solutions. A model of Return on CPS (RoCPS) is developed specifically to quantify the potential profits by CPS applications, which is defined in Formula (1) with reference to the method of Return on Investment [13].

$$\mathrm{RoCPS} = \sum_{i=0}^{n} \frac{p_i}{(1+r_i)^i} \Big/ \sum_{i=0}^{n} \frac{c_i}{(1+r_i)^i} \tag{1}$$

p_i is the total profit in period i, c_i is the total cost in period i, r_i is the discount rate in the period i, n is number of periods. The value of RoCPS indicates how much return can be gained after adopting CPS compared with its implementation costs. In the matching platform, a universal template sheet for cost and profit estimation is provided, in which all cost drivers and benefit types are identified and monetization rules are given, thus facilitating non-experts to make assessment (see more in [14] about our research work).

4 Application of the CPS Matching Platform

The developed matching platform was applied to conceptualize a CPS solution at a small-sized German EV manufacturer. The SME wishes to track the order status during the end-of-line assembly by a wireless identification and localization of the initial chassis place and the final mounting place, in order to bring more transparency in intralogistics.

Firstly, the production manager of the SME chose the use case type on the matching platform. Following information processes of the use case were suggested by the platform: identification, localization, raw data transmission, raw data processing, information transmission and data storage. Among them the identification, localization and data transmission are the mayor challenging selection task to construct the CPS.

Secondly, the matching platform asked the production manager to fill out a requirement catalog for each information process. The requirement specifications in the catalog are derived from the evaluation scheme introduced in 3.1. For this tracking and tracing

case, the resolution of the localization within less than one meter, the data rate larger than 1 Mbit/s and a scalable IT-infrastructure are some most decisive requirements. Thirdly, the matching platform automatically filtered the suitable solutions, because all CPS technology specifications stored in the platform have been assessed in advance with the identic evaluation scheme as the requirement specifications (see Figure 2). In analogy to the morphological box [11], the platform can also generate CPS technology combinations, which fulfills the requirements of all information processes. Now the production manager can either take the optimal combination suggested by the platform or combine the technologies by himself and compare the functional fulfillment based on the grading.
Finally, the RoCPS was calculated with the given calculation sheet and the profitability of CPS concepts was evaluated. Figure 2 shows that the first CPS concept suitable for the use case consists of RTLS and LTE; the second concept suggests 5G as the transmission technology. Since the use of mobile communication systems in production requires a base station and customized networks, which is an expensive option, the SME decided to choose Bluetooth connections via Beacons by self-combination.

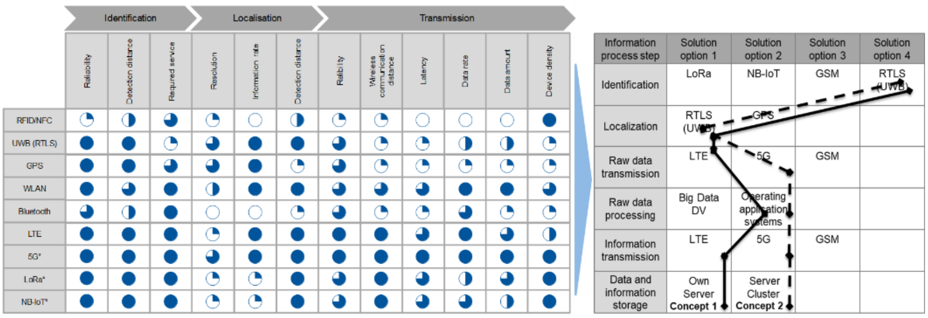


Figure 2. Two CPS concepts suggested by the matching platform for the use case

5 Summary and Outlook

This paper developed a matching platform for SMEs to select appropriate CPS solutions from both functional and economical perspectives. It was demonstrated by a small-sized German EV manufacturer, supporting them in CPS selection for tracking and tracing during the end-of-line assembly. The matching platform bridges the gap between inexperienced users and CPS suppliers, which can promote the digitalization transformation of SMEs. In future, it is expected to be optimized by more uses of public.

6 Acknowledgements

The research has been funded by the Leitmarkt IKT.NRW program of the European Federation for Regional Development (EFRE) within the scope of the project “cyberKMU²”. The authors acknowledge the EFRE for their support.

References

1. Lee, E.A.: Cyber Physical Systems: Design Challenges. In: 11th IEEE symposium: object oriented real-time distributed computing (ISORC), Orlando, FL, pp. 363-369 (May 2008)
2. Geisberger, E., Broy, M.: Living in a networked world: Integrated research agenda Cyber-Physical Systems (agenda CPS). 1st ed., Herbert Utz Verlag, München (2015)
3. Schleipen, M.: Glossar Industrie 4.0 des Fachausschuss VDI/VDE-GMA 7.21 „Industrie 4.0". Fraunhofer IOSB (2015)
4. Stich, V., Deindl, M., Jordan, F., Maecker, L., Weber, F.: Studie – Cyber physical Systems in der Produktionspraxis. Wuppertal (2015)
5. China National Science Foundation.: Cyber-Physical Systems (CPS). White paper report (2016)
6. Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte, VDI 2221 (1993)
7. Technisch-wirtschaftliches Konstruieren- Vereinfachte Kostenermittlung, VDI 2225-1 (1997)
8. Technisch-wirtschaftliches Konstruieren-Technisch-wirtschaftliche Bewertung, VDI 2225-3 (1998)
9. Zangemeister, C.: Nutzwertanalyse in der Systemtechnik: Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen. 4th ed. München: Wittemannsche Buchhandlung (1970)
10. Zangemeister, C.: Nutzwertanalyse in der Systemtechnik: Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen. Teilw. zugl.: Berlin, Univ., Diss., 1970, 5th ed. Norderstedt: Books on Demand (2014)
11. Jordan, F., Bernardy, A., Stroh, M., Horeis, J., Stitch, V.: Requirements-based matching approach to configurate Cyber-Physical Systems for SMEs. In: PICMET '17: Technology Management for Interconnected World, Portland, OR (2017)
12. Bernardy, A., Jordan, F., Schuh, G., Stich, V., Zeller, V.: Basic Methodology for Cyber Physical System Modelling. In: PICMET'18: Managing Technological Entrepreneurship: The Engine for Economic Growth, Honolulu, HI, (2018)
13. Hering, E.: Investitions- und Wirtschaftlichkeitsrechnung für Ingenieure (essentials). Springer-Verlag, Berlin Heidelberg (2014)
14. Burggräf, P., Dannapfel, M., Bertling, M., Xu, T.: Return on CPS (RoCPS) – An Evaluation Model to Assess the Cost Effectiveness of Cyber-Physical Systems for Small and Medium-Sized Enterprises. In: PICMET'18: Managing Technological Entrepreneurship: The Engine for Economic Growth, Honolulu, HI (2018)

Track 2:

Logistic Analytics

Track Chairs:

Dr. Dominik Kreß
Universität Siegen

Dr. Alena Otto
Universität Siegen

Prof. Dr. Erwin Pesch
Universität Siegen

An Empirical Study of Customers' Behavioral Intention to Use Ridepooling Services – An Extension of the Technology Acceptance Model

Marc-Oliver Sonneberg[1], Oliver Werth[1], Max Leyerer[1], Wiebke Wille[1], Marvin Jarlik[1], and Michael H. Breitner[1]

[1] Leibniz University Hannover, Information Systems Institute, Hannover, Germany
{sonneberg,werth,leyerer,wille,jarlik,breitner}
@iwi.uni-hannover.de

Abstract. Shared mobility services for passenger transportation become increasingly popular all over the world. As services like carsharing are already well-established and well-accepted, ridepooling services are at their early stage and currently within first implementations. The most critical success factor of such services is the customer acceptance. We investigate the acceptance of 115 German questionnaire respondents using and extending the Technology Acceptance Model. Results indicate that the success factors of the developed model serve as useful predictors of the behavioral intention to use ridepooling services. Perceived compatibility was identified to have the strongest impact whereas perceived ease of use and perceived safety are not relevant for accepting ridepooling services. Based on these findings, our paper provides management implications and recommendations to improve acceptance and success of ridepooling services in Germany.

Keywords: Ridepooling, Passenger Transportation, Urban Mobility, Technology Acceptance Model, Structural Equation Modeling.

1 Introduction and Motivation

Urban areas are confronted with a multitude of challenges as high emissions, poor air quality, fossil fuel dependency, traffic volume, and congestion [1]. With increasing consciousness for sustainability and environmental responsibility, the need for innovative solutions tackling these problems is emerging. As a consequence, the sharing economy has been arisen from the idea that sharing a good or a service is often more advantageous than owning it as resource inefficiencies are reduced. Regarding passenger transportation, rideservices depict a possibility for individuals to share a car or a trip in different modes. Supported by technological developments and the digitalization of processes, companies are able to offer reliable modes of dynamic on-demand rideservices; concurrently, customers can participate easily through the use of immediate communication with connected mobile devices [1-3]. In this way, the

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

information system (IS) domain can be characterized as enabler for digital economies and corresponding digital services as well as business models [4].

The most recent development of digitally supported rideservices is ridepooling (also referred to as shared ridehailing). Using big (geo) data analytics approaches together with intelligent algorithms, passengers are aggregated to groups and allocated to the best option of available vehicles in real-time and on-demand. Thereby, multiple passengers share a ride in the same vehicle to increase transport efficiency. Recent studies from New York City demonstrate ridepooling's efficiency, as it was shown that the traffic volume can be significantly reduced when using high-capacity ridepooling instead of individual taxi services [2, 5]. Due to this positive impact, ridepooling becomes of increasingly interest to cities worldwide. In Germany, potential ridepooling providers are planning to offer this service in order to take advantage of the untapped potential. Except of overcoming the legal barriers, the acceptance of ridepooling services is a critical factor for being successful in the long term. It represents the first step in the adaption process and results in actual usage. Especially in the early stages of innovation development, its investigation is of high relevance because a modification of the service is still possible based on changed customers' needs and requirements [6].

In the literature, the acceptance of innovations has been already investigated by different models regarding various contexts [7]. With the help of hypothesis-testing studies, pertinent factors for these acceptance models have been identified. The most used and established model is the Technology Acceptance Model (TAM) [8]. With regard to the ridepooling context neither a model nor success factors have been studied yet. To address this research gap, we investigate the customer acceptance using TAM as basic theory and extending it in order to fit the research field of ridepooling services. The following research question guides our examination:

RQ: Which constructs influence the customer acceptance of ridepooling?

To answer this question, the article is structured as follows: first, the background is explained containing urban mobility, TAM, and related literature. The methodology covering the hypothesis development, the study description, and the results build the third section. In the fourth section, obtained results are discussed and the contributions of our approach are highlighted. Recommendations, limitations, and further research are elaborated in the fifth section. At last, we complete our article with conclusions.

2 Research Background

2.1 Urban Mobility, New Mobility Services, and Ridepooling

As an outcome of numerous individuals' decisions, urban mobility describes passenger movements within the city environment. Regarding the travel behavior, passengers decide upon vehicle ownership, individual or collective transportation, and mode of transport (e.g., car, bike, feet, tram, etc.). In contrast to private vehicle ownership, people are able to choose Mobility-as-a-Service (MaaS) (e.g., bus, taxi, carsharing, bikesharing, mass transit, etc.) to carry out daily activities. The individuals' decisions are affected by several key factors which are dynamic and interacting [9]:

- *demographic trends* – population (e.g., age, growth, density), licensed drivers;
- *transportation options* – private vehicles, carsharing, mass transit, taxi;
- *infrastructure* – road network, traffic management systems;
- *user preferences* – social mobility preferences, residence;
- *transportation costs* – fuel, transit, parking, ownership costs; and
- *macroeconomic facts* – economic growth, global warming, employment, pollution.

Besides these factors, technological developments influence the way people move, as they address economic, ecological, and societal problems within people's environment [10]. This results in new mobility concepts, respectively new mobility services (NMS) as subcategory of MaaS. Emerging technologies – such as digitalization, high-speed computing, location data, accurate sensors, wireless connectivity, social media expansion, and new usage-based pricing schemes – serve as enabler for NMS [9]. Thereby, NMS represent potential solutions to allow for a more convenient, efficient, and flexible transportation for different individual purposes of travelling. NMS contribute to a mobility evolution because they incrementally change the travel behavior towards a multimodal and less car-centric system, particularly in urban areas. They dissolve the boundaries of what is owned and shared [9]. Present trends take a stronger shift towards sharing with a special focus on sustainability [11]. As the use of shared transportation modes partially substitutes private car ownership, the number of cars on and off the road can be reduced. Consequently, traffic density, travel costs and time, fuel consumption per person, and air pollution are reduced [12, 13]. One reason for choosing NMS instead of owning a car is the problem of missing parking space in urban areas. Compared to public transportation, NMS have the advantage of high flexibility, especially regarding point-to-point service. In addition, short waiting times and easy payment methods are further benefits [14]. The following list gives an overview of the latest NMS including a brief explanation [15]:

- *Carsharing*: Users pay money based on the used time or distance for renting a car (variants: business-to-consumer or consumer-to-consumer; station-based or free-floating).
- *Carpooling* or *Ridesharing*: A vehicle is used by individuals who take a ride at the same time in the same car from and to similar destinations. The matching is done by an intermediary company or by an informal system of the users.
- *Ridehailing*: Determining the trip's start and end point, a passenger demands a transport service offered by companies or individuals.
- *Ridepooling* or *Shared Ridehaling*: Users hail a shuttle to designated pick-up points near their location. Passengers with similar routes are matched and transported together in one vehicle.

The most recent development regarding NMS is the shared travelling respectively pooling of users with overlapping routes. Thus, customers share the costs of the trip resulting in prices between taxi and mass transit charges. Exemplary services are *uberPOOL*, *Lyft Line*, and *MOIA*. The concept of ridepooling comprises that users hail shuttles via an app and get on the shuttle at designated pick-up points near their location. An algorithm optimizes each vehicle's route in terms of travel time and capacity to enable shared trips with overlapping routes. As a result, the number of cars and thereby road traffic's negative impacts can be reduced significantly. Dealing with the urban mobility in an American context, a study by Alonso-Mora et al. [2] predicts that 99% of the taxi demand in New York City could be served by 25% of the utilized vehicles if using high capacity ridepooling. Due to the economic, environmental, and societal advantages, ridepooling offers a high potential to different target groups. Besides competition and political issues, an important barrier to overcome appears to be the acceptance and the conclusive usage by the population.

2.2 Technology Acceptance Model

Consumer acceptance can be defined as the "relatively enduring cognitive and affective perceptual orientation of an individual" [8]. Thereby, the acceptance process of individuals depends on the tradeoff between benefit and effort of using an innovation or technology [16]. Investigating these psychology processes causing special human behavior is complex and difficult [17]. For solving this problem, lots of social-psychological models have been developed in the last decades to explain and predict technology acceptance as well as usage of individuals [18].

In scientific literature, the most popular cited model for that is TAM [8]. TAM is an adaption of the Theory of Reasoned Action (TRA) which was originally developed by Fishbein and Ajzen [19] to predict human behavior [20]. Davis [21] and Davis et al. [22] adopted TRA to the acceptance research in IS contexts and used it as a basic theory to explain the relationship between the individual's reaction of using a technology, the intention, and the actual usage of it [23]. For that it utilizes behavioral intention to use (BI) to predict actual behavior and focuses on the identification of relevant factors for adopting an innovation or technology [24]. TAM can also be transformed into a measurement of customer acceptance in other varieties of settings and technologies like internet banking, mobile service, online tax service, or teacher's technology usage which has been investigated by several researchers (e.g., [6], [20], [25]). Both models, TRA and TAM, are based on individual beliefs which determine and affect its attitude towards a technology in a given situation [26]. Beliefs are defined as "the person's subjective probability that performing the target behavior will result in salient consequence" [21]. These beliefs are the internal psychological variables, for instance, the individuals' characteristics in the models. They function as mediators of all external variables like individuals' characteristics which also may affect the usage of an innovation [27]. Therefore, they have an indirect effect on the BI [21]. In TAM, these beliefs consist of the two factors perceived usefulness (PU) and perceived ease of use (PEOU) which have a relationship to each other. At TRA, attitude towards use (ATU) and subjective norm (SN) represent the beliefs [28]. In addition, TAM represents the motivational variables which lead to the actual system usage. They reflect a tendency which is built directly at the beginning of being in contact with the innovation. This enables researchers to test the innovation in an early stage [21]. Because of the mentioned descriptions, we make use of TAM in combination with constructs of the TRA in order to measure the acceptance factors of ridepooling.

2.3 Related Research

Chowdhury and Ceder [29] provide a literature overview of public transport studies analyzing willingness to ride as well as the related acceptance. Investigations on the acceptance of NMS are not contained within the review. However, there are some studies that deal with the different carsharing variants and their acceptance. Ohta et al. [30] present an article that examines the acceptance of carsharing in Japan. As result, the attitude to conduct carsharing for car-owners was low and the BI was quite high for non-owners. People living in urban areas had a greater acceptance for carsharing than rural inhabitants. Dütschke and Peters [31] conduct an empirical analysis on sustainable modes of transport like carsharing and electric vehicles. As a result, perceived compatibility (PC) was the most influencing factor towards ATU carsharing across all sociodemographic groups. Fleury et al. [32] examine the acceptance of corporate carsharing in France based on the Unified Theory of Acceptance and Use of Technology (UTAUT) framework. The most important dimension in determining behavioral intentions about corporate carsharing was effort expectancy. Besides, the added perceived environmental friendliness had only a small effect on behavioral intentions, mediated by performance expectancy. Another quantitative study by Efthymiou et al. [33] investigate the factors affecting the adoption of carsharing systems by young drivers. Further, Cheng et al. [34] examine the motivation of users to participate on a carsharing platform.

Regarding ridesharing, Giang et al. [26] investigate the customers' BI such services. The authors ascertained the positive effect of ATU ridesharing applications on BI them based on a Vietnamese study. Results demonstrate that PU and PEOU had positive influence on attitude toward ridesharing behaviors. The constructs ATU, SN, and perceived behavior control further played critical roles in predicting the BI ridesharing applications. Another article by Wang et al. [35] investigate the customers' BI such services extending TAM on the new constructs personal innovativeness, environmental awareness, and perceived risk (PR). Results demonstrate that these constructs were positively associated with customers' BI ridesharing services. On the other hand, PR was negatively associated with the BI as well as PU. Further, personal innovativeness was negatively related to PR. Other quantitative analyses are for instance Delhomme and Gheorghiu [36], who conduct a French study of users and non-users, or Wright et al. [37], who focus on the acceptance of a ridesharing-platform.

Concerning acceptance analyses on ridehailing or ridepooling, literature lacks using common applications of theories like TAM, TRA, or others. To the best of our knowledge, no article focuses on the acceptance of these services itself. There exist only short surveys revealing that around 79% of the ridehailing-users would use ridepooling, depending on factors like costs and number of passengers extracted from a Brazilian study [38]. Clewlow and Mishra [15] investigate the adoption and use of ridehailing services in San Francisco. Studies on the acceptance of apps or platforms for ridehailing services are more frequent, as for instance, Tan et al. [39] or Ruangkanjanases and Chayanee [40]. For ridepooling service apps or platforms, no scientific studies are existent. Thus, ridepooling is a quite new concept which currently starts being implemented and has not been addressed in acceptance studies so far.

3 Methodology

3.1 Hypothesis Development

One of the goals of TAM is the quantification of the influence of behavioral intentions on actual systems usage. Measuring the actual system usage is in most contexts difficult, especially if the acceptance of a system is investigated which is still at its early stage of implementation, such as ridepooling [41]. The actual system usage is influenced by the BI a technology. It measures the strength "of one's willingness to exert effort while performing certain behaviors" [41]. By a high degree of accuracy, the actual system usage can be very well approximated by behavioral intentions which a lot of studies already found out in different contexts [17]. According to the TRA, behavioral intention is the direct factor of the appropriate behavior [25]. Furthermore, it is more suitable "for a survey-based research", because beliefs can be measured at the same time [27]. Based on this relationship, we choose the BI ridepooling as the dependent variable in our approach. BI is directly influenced by ATU. It describes "the degree to which using a technology is positively or negatively valued by an individual" [42]. According to the NMS context, Giang et al. [26] explore the positive effect of ATU ridesharing applications on BI. In view of these findings regarding ridepooling services it can be hypothesized:

H1: ATU has a significant positive effect on the BI ridepooling services.

ATU does not only represent the main antecedent of BI but also functions as the key mediator between it and the other influencing factors [43]. According to the TRA, attitude will be developed by individual beliefs. These beliefs arise through learning processes and consequently affect attitudes [27]. Therefore, a better understanding of ridepooling may lead to a superior ATU ridepooling and consequently to a higher acceptance. These beliefs are represented by the constructs of PU and PEOU. Therefore, the following hypotheses can be derived:

H2: PU has a significant positive effect on ATU ridepooling services.

H3: PU has a significant positive effect on BI ridepooling services.

According to TAM, PEOU is another driver of ATU and represents also a crucial belief [7]. This term addresses the complexity degree of an innovation "to which (…) [it] is perceived as relatively difficult to understand and use" [16]. In a ridepooling context, the mobile application to order the service should work without errors, its functions should be instinctively understandable, and the user should be able to choose between different payment methods. Besides, the amount of effort must be in a proportionate relation to the usage. Therefore, the following hypothesis can be developed:

H4: PEOU has a significant positive effect on ATU ridepooling services.

Apart from the direct effect of PEOU on ATU, Davis [16] found the indirect effect via PU. The easier the innovation is perceived to be, the higher seems its usefulness in the case that more effort can be put into other activities. Therefore, the advantage of PU on ATU would be weakened by the uneasiness to use [16]. For this reason, the following hypothesis can be derived:

H5: PEOU has a significant positive effect on PU of ridepooling services.

According to the diffusion theory of Rogers [44] several factors are responsible for the adaptation of innovations. Being quite similar to the belief concept of TAM, it proposes PC as one of the core factors for customer acceptance of new technologies [45]. It can be defined as “the degree to which an innovation is perceived as being consistent with the existing values, needs and past experiences of the potential adopters” [41]. In the literature, lots of studies discovered the high importance of compatibility for measuring and explaining technology acceptance behavior. For example, Schierz et al. [8] already found the positive effect of PC on ATU, PU, and BI a technology. Based on these studies the following hypotheses can be developed:

H6: PC has a significant positive effect on the PU of ridepooling services.

H7: PC has a significant positive effect on the ATU ridepooling services.

H8: PC has a significant positive effect on the BI ridepooling services.

Apart from reaching benefits of using new technologies (e.g., PU and PEOU), it is still hard to evaluate the unknown [8]. As new technologies are uncertain in their outcomes, they seem to be riskier for customers [46]. Especially when it comes to transportation modes, feelings of safety are associated with PR that is why the selection of transportation modes is dependent on perceived safety (PS). That is why PS is a relevant issue in the car context regarding the acceptance which Osswald et al. [6] confirmed in their study. They defined this construct “as the degree to which an individual believes that using a system will affect his or her well-being” [6]. It addresses the capability of an individual to estimate the situation as being dangerous or safe. Regarding to the context of ridepooling services, the element of safety exemplarily includes the estimation of the driver’s ability and trustworthiness [6]. Another safety problem might be the traveling with unfamiliar passengers [29]. Therefore, we construct the following hypothesis:

H9: PS has a significant positive effect on the ATU ridepooling services.

Especially at the beginning of a diffusion or development of an innovation, the social context plays a major role within the adaptation process. The influence of the social context is represented by the factor SN which can be “defined person’s perception that most people who are important to him think he should or should not perform the behavior in question” [19]. Already Davis et al. [22] suggested to further research the relationship between TAM and the effect of social influences on usage behavior. That is why the following hypothesis can be set:

H10: SN has a significant positive effect on the BI ridepooling services.

3.2 Description of Our Study

Primary data was collected through an online-survey. We shared our questionnaire to German participants only. Therefore, the complete questionnaire was translated into German before. To generate high response rates in different areas, the survey was spread over social media, forums, and via e-mail using our existing network on transportation research. We conducted our examination in the first half of March 2018.

Our questionnaire begins with a short definition of ridepooling services to ensure similar knowledge of the respondents. The measurement items base on different applications and were modified to suit the research context of ridepooling services. We adapted the items for the construct ATU from Cheng et al. [34] and Delhomme and Gheorghui [36]. Items for PU and PEOU were adapted from Davis [16]. SN was adapted from Madigan et al. [18]. Items for the constructs PC and BI were taken and adapted from Schierz et al. [8]. As described, we added the construct PS to our model, adapting the items from the study of Osswald et al. [6]. For instance, we asked questions about concerns with unfamiliar passengers or trust into the driver of a ridepooling vehicle. For measuring the items, a seven-point Likert scale was used, ranging from 1 ("strongly disagree") to 7 ("strongly agree"). The underlying codebook is presented in Table 1.

Regarding demographics, 115 respondents took part in the survey, whereas 48 (42%) were female and 66 (57%) were male. Regarding the age, 94 participants responded to the group of 20-29 years old, representing 70% of our total data set, 15 people (13%) were between 30-39 years. In terms of their profession, 48 were students (42%) and 48 employees (42%). We asked some general questions about the usage of ridepooling services in the past and relevant factors for the future. The additional questions could be answered on a multiple selection with an open text field. 77 (67%) of the participants have already heard of ridepooling services. 91 out of 115 people surveyed (79%) did not used ridepooling services in the past. We asked also about the (possible) purposes to use a ridepooling services. Most frequently, participants would use ridepooling to drive to a train station or airport (mentioned 74 times) and to drive to a nightclub in their leisure time (mentioned 70 times).

Table 1. Presentation of underlying Codebook

Constructs	*Items*	*Constructs*	*Items*
ATU 1	More comfortable mode of transportation	PC 1	Usage depends on transportation possibilities
ATU 2	Independence of movement	PC 2	Usage fits lifestyle
ATU 3	Pleasant service	PC 3	Preferring of ridepooling services
ATU 4	Advantageousness of usage	PEOU 1	Easy execution of the steps to usage
ATU 5	Good idea to use	PEOU 2	Precise and comprehensible interaction
BI 1	Usage in contrast to other transportation modes	PEOU 3	Easy usage
BI 2	Probability of usage in the future	PEOU 4	Learning would be easy
BI 3	Willingness of usage in the future	PS 1	Dangerous usage
BI 4	Intention of usage in the future	PS 2	Ride with unfamiliar passengers
SN 1	Importance by family and friends	PS 3	Trust in the driver
SN 2	Good idea by family and fiends	PU 1	Useful mode of transportation
SN 3	Recommendation by family and friends	PU 2	Simplification of the transport
		PU 3	Fast transport possible
SN 4	Usage by family and friends	PU 4	Improvement of the transportation selection
Notes:	ATU: Attitude towards Use; BI: Behavioral Intention to Use; SN: Subjective Norm; PC: Perceived Compatibility; PEOU: Perceived Ease of Use; PS: Perceived Safety; PU: Perceived Usefulness		

3.3 Results

We performed Structural Equation Modeling with the help of the software package *SmartPLS* (version 3.2.7). For validating the reflective measurement models the indicator reliability must be given. This is proven by a factor analysis which tests if more than 50% of the variance of an indicator can be traced back to the associated latent variable. The indicator can remain in the dataset if the factor loadings are above 0.5 [47]. The factor analysis show factor loadings above the critical value for all indicators except of two. In consequence, we dropped SN 4 and PC 2 for our investigations. Our scale reliability tests found Cronbach´s alpha values greater or equal 0.7. Therefore, we found an acceptable internal consistency [48]. Table 2 shows the mean values, the standard deviations (Std. Dev.), and the discriminant validity analysis of our model. We compared the roots square of average variance extracted (AVE) values and the correlations between the constructs [49]. The diagonal elements in bold face in Table 2 represent the square root of AVE values. All of them are larger than the correlations among constructs. In result, convergent and discriminant validity are verified [50].

Table 2. Mean Values, Std. Dev., and Discriminant Validity of the Structural Equation Model

Construct	Mean	Std. Dev.	ATU	BI	PC	PEOU	PS	PU	SN
ATU	4.73	1.40	**.899**						
BI	4.41	1.69	.800	**.894**					
PC	3.93	1.63	.747	.814	**.905**				
PEOU	5.58	1.27	.453	.477	.518	**.904**			
PS	5.57	1.16	.428	.479	.411	.394	**.872**		
PU	4.27	1.45	.841	.783	.792	.475	.401	**.862**	
SN	4.70	1.37	.697	.664	.566	.441	.403	.662	**.951**

In support of H1, there is a positive significant, but moderate influence of ATU ridepooling services on BI them (β=0.29; p<0.001). Moreover, the path coefficient of β=0.65, (p<0.001), points to a strong positive relationship between PU of ridepooling services and the ATU them. Thus, H2 can be affirmed. Regarding H3, even though the structural link from PU to BI ridepooling services is still positive (β=0.09; $p\geq$0.1), no significant relationship can be proved. In addition, the results do not provide significant evidence for the effects of PEOU of ridepooling services on ATU them (β= 0.01; $p\geq$0.1) and on PU (β=0.09; $p\geq$0.1). As result, H4 and H5 are not supported. The strongest significant influence is found at PC of ridepooling services on PU of them with a beta of 0.75 and a p-value<0.001, as hypothesized and supported in H6. Furthermore, a positive influence reveals PC on ATU ridepooling services (β=0.19; p<0.01) so that H7 can also be supported. As hypothesized in H8, with a path coefficient of β=0.44, PC has a positive impact on BI ridepooling services (p<0.001) and is therefore supported. The hypothesized significant effect of ridepooling services' PS on the ATU them was found to be nonsignificant (β=0.08; $p\geq$0.1), thus H9 is rejected. The hypothesized assumption of the influence of SN on BI ridepooling services (H10) can be confirmed to be significant (β=0.16; p<0.1). Due to these results, H1, H2, H6, H7, H8, and H10 can be accepted, whereas H3, H4, H5, and H9 need to be rejected.

Apart from investigating the path coefficients, the quality of the endogenous variables must be checked as well. The construct of PU of ridepooling services is explained by its exogenous variables of adjusted $R^2= 0.63$. ATU ridepooling services as well as BI them are quite similar: the former shows an explanatory power of adjusted $R^2=0.72$; the second of adjusted $R^2=0.75$. With regard to the definition, these results are satisfactory. Figure 1 summarizes the results of our path analysis presented before.

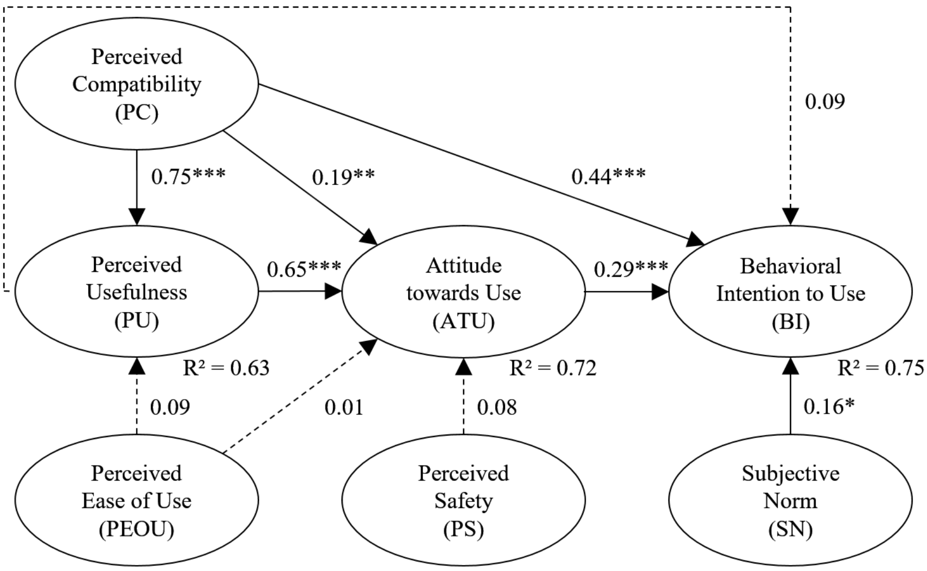


Figure 1. Results of the Path Analysis
*Notes:*p<0.1; **p<0.01; ***p<0.001; n=115; Dotted lines represent insignificant paths*

4 Discussion and Contributions

The insignificant direct influence of PU on BI (H3) is quite unexpected as this relationship has been confirmed in the literature many times in other contexts. For instance, Wang et al. [35] who analyzes the customers' BI ridesharing found a significant positive relationship between PU on BI. Regarding the ridesharing context, the service is seen as useful and has therefore, in contrast to ridepooling, an influence on BI. PU represents a fundamental driver of BI and counts therefore to the key constructs of the TAM [51]. However, a possible reason could be that the majority of respondents have not yet used ridepooling services. Under certain circumstances, it may be not the usefulness but the experience and its accompanying pleasure that is a reason for participating in such a concept. The journey in a modern and appealing vehicle with a slight cost disadvantage compared to conventional mass transport could be an explanation for this. According to these results, the direct relationship cannot be transferred to the ridepooling context with the empirical data of our study. Consequently, the mentioned direct effect of beliefs on BI ridepooling services cannot be affirmed. In addition, Lee et al. [41] found ten non-significant and even 17 not applicable relationships of in total 101 investigated studies in their literature review. This refers to the fact that there are also other studies with the same inexplicable results.

In contrary to that, the role of PEOU in the TAM is already critically discussed in the literature. Many researchers question this construct as a fixed component in the TAM due to the existence of many controversy studies regarding the unstable measurement of PEOU. Therefore, rejecting H4 and H5 is not quite unexpected. PEOU of ridepooling has no significant influence on PU and ATU of ridepooling. In contrast to our results, Giang et al. [26] found a significant relationship of PEOU on ATU in a ridesharing contexts. Lee et al. [41] summarized that there are 13 non-significant and 19 not applicable relationships between PEOU and PU as well as 24 non-significant and 19 not applicable relationships between PEOU and BI. This result already shows the doubtfulness of this construct [41]. According to Subramanian [52] the low or not existing importance of PEOU on the acceptance process arises at innovations which are intuitively easy to use, so this factor has no impact on this process any more. This argumentation comes along with the mean value of 5.58 for PEOU (see Table 2) which is the highest of all constructs and therefore an indicator for the easiness to use ridepooling services. Moreover, literature often mentions that the impact of PU on ATU is stronger than PEOU on ATU. This points out the subordinate role of PEOU in the TAM [6, 26]. This can also be confirmed in our study as PU is the more important construct for predicting the ATU ridepooling services. Therefore, the construct of PEOU plays a subordinate role for explaining the acceptance of ridepooling services.

By rejecting H9, it was shown that the introduced variable of PS is not relevant for the acceptance process, or to be more precise on the ATU in a ridepooling context. A possible reason for this not significant relationship is that the journey with instructed drivers or other passengers is not considered unsafe as shared transportation means or MaaS are already used and therefore has no influence on the acceptance of ridepooling. This stands in contrast to the findings of Agatz et al. [53], who state that factors as security and safety must be ensured.

Concluding, the model and its connections can be largely confirmed. Most of the extended TAM variables contribute to the prediction of BI ridepooling services. As a core variable of TRA, SN is identified as having a low influence on BI ridepooling services and therefore contributes to a better prediction of its acceptance. This is in line with the findings of Giang et al. [26] who found a significant positive relationship between SN and BI. As already suggested, this extension should be a core variable in the model, in particular regarding innovations with unexperienced users. In addition, the high relevance of PC in the context of ridepooling services has been identified. Thus, for accepting ridepooling services, the technology and service itself must be compatible to the individual lifestyle. These strong effects observed in the data suggest to include this construct as a permanent factor in the TAM regarding the context of ridepooling services. Also, the influence of PEOU and PU deviated from expectations in some cases because the data was too inconsistent for accepting H3, H4, and H5. As a result, the suggestion of the available theory cannot be totally confirmed by the empirical data. This gives cause to rethink their roles in the model especially in the context of ridepooling services.

5 Recommendations, Limitations, and Future Research

The results indicate a stronger preferred use of ridepooling services for leisure activities, while daily commuter trips are not the main objective of such services. Respondents do not seem to have any problems regarding the handling of the service, as they evaluated PEOU very high. Scientific research suggests, that perceived behavioral control has a positive influence on the adoption of innovations [54]. As the general acceptance of ridepooling services is existent, providers must ensure the availability and reliability of the booking process as well as the service itself. Therefore, the level of travel costs as well as the use of ecologically friendly vehicles play a crucial role for the success of a ridepooling service provider.

Since ridepooling is, especially in Germany, a rather new mobility concept, relatively little information for customers about this concept exists. With greater awareness and more practical implementations, ridepooling's acceptance might be different. As our study focuses on Germany, comparative studies across other countries or cultures can be conducted to identify similarities or differences. In order to achieve better predictive results, the contextual factors and external variables as for example gender, income, system characteristics, design features, and personality traits which are neglected in our research can be also included. According to Davis [16], they have a direct effect on PU and PEOU. In addition, some of the correlations among our constructs are very high as presented in Table 2, e.g., PU and ATU with a correlation of 0.841. Possible explanation of this could be a lacking differentiation of the constructs and the corresponding items. To counteract this points, future research can base on other items or other theories, as for instance UTAUT by Venkatesh et al. [23], since moderating variables are included in the analysis. Future research can also replicate the same study after a period of time in order to identify differences caused by more ridepooling providers or more experience with ridepooling services. Thus, a longitudinal study is recommended, ideally with an evenly distributed sample of participants. In this way, it can be measured whether potential customers will accept, adopt, and actually use ridepooling services.

6 Conclusions

Our paper gives insights regarding the customer acceptance of ridepooling services which support the society's shift towards a digital and shared economy using various IS solutions. As this field of research is still underdeveloped in the literature, our study addresses this research gap to satisfy the increasing interest and importance of this topic in practice. The TAM with its core determinants was used in combination with constructs of the TRA. To fit the research context of ridepooling services, it was extended by the constructs PC and PS. The results of the analysis show that BI ridepooling services is well explained by the developed model. Unexpectedly, the relationship between PU of ridepooling services as well as the BI them plays only a subordinate role. In addition, the results do not provide significant evidence for the impact of PEOU and PS on ATU ridepooling services. In contrary to that, the high relevance of PC for predicting the acceptance of ridepooling services has been identified as it has the most significant impact of BI them. This finding shows that the acceptance process of ridepooling services has already started. The trend of ridepooling is getting adopted by Germans as well. Therefore, companies must tackle this challenge to exploit the high potential this market offers. Analysis have revealed that companies can increase their acceptance, for instance, using electric vehicles or offering low costs.

References

1. Chan, N.D., Shaheen, S.A.: Ridesharing in North America: Past, Present, and Future. Transportation Reviews 32(1), 93-112 (2012)
2. Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D.: On-Demand High-Capacity Ride-Sharing via Dynamic Trip-Vehicle Assignment. In: Proceedings of the National Academy of Sciences, 114(3), pp. 462-467 (2017)
3. Farin, N.J., Rimon, M.N.A.A., Momen, S., Uddin, M.S., Mansoor, N.: A Framework for Dynamic Vehicle Pooling and Ride-Sharing System. In: IEEE International Workshop on Computational Intelligence, pp. 204-208 (2016)
4. Barrett, M., Davidson, E., Prabhu, J., Vargo, S.L.: Service Innovation in the Digital Age: Key Contributions and Future Directions. MISQ 39(1), 135-154 (2015)
5. Nicolas, L., Moura, S.J.: Optimal Routing and Charging of Electric Ride-Pooling Vehicles in Urban Networks. Working Paper: UC Berkeley: Energy, Controls, and Applications Lab (2016)
6. Osswald, S., Wurhofer, D., Trösterer, S., Beck, E., Tscheligi, M.: Predicting Information Technology Usage in the Car: Towards a Car Technology Acceptance Model. In: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 51-58. ACM, New York, (2012)
7. Karahanna, E., Agarwal, R., Angst, C.M.: Reconceptualizing Compatibility Beliefs in Technology Acceptance Research. MISQ 30(4), 781-804 (2006)
8. Schierz, P.G., Schilke, O., Wirtz, B.W.: Understanding Consumer Acceptance of Mobile Payment Services: An Empirical Analysis. Electronic Commerce Research and Applications 9(3), 209-216 (2010)
9. Spulber, A., Dennis, E.P., Wallace, R., Schultz, M.: The Impact of New Mobility Services on the Automotive Industry. Center for Automotive Research (CAR), Ann Arbor (2016)
10. Proff, H., Fojcik, T.M.: Mobilität und Digitale Transformation - Technische und betriebswirtschaftliche Aspekte - Einordnung. In: Proff, H., Fojcik, T. (eds.), Mobilität und Digitale Transformation, pp. 1-5. Springer Gabler, Wiesbaden (2018)
11. Cohen, B., Kietzmann, J.: Ride On! Mobility Business Models for the Sharing Economy. Organization & Environment 27(3), 279-296 (2014)
12. Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.E., Wang, X., Koenig, S.: Ridesharing: The State-of-the-Art and Future Directions. Transportation Research Part B 57, 28-46 (2013)
13. Kuntzky, K., Wittke, S., Herrmann, C.: Car and Ride Sharing Concept as a Product Service System – Simulation as a Tool to Reduce Environmental Impacts. In: Shimomura, Y., Kimita, K. (eds.): The Philosopher's Stone for Sustainability, pp. 381-386. Springer, Berlin / Heidelberg (2013)
14. Rayle, L., Dai, D., Chan N., Cervero, R., Shaheen, S.: Just a Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. Transport Policy 45, 168-178 (2016)

15. Clewlow, R.R., Mishra, G.S.: Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States. Institute of Transportation Studies, University of California, Davis, Research Report UCD-ITS-RR-17-07 (2017)
16. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MISQ 13(3), 319-340 (1989)
17. Ajzen, I.: The Theory of Planned Behavior. Organization Behavior and Human Decision Processes 50(2), 179-211 (1991)
18. Madigan, R., Louw, T., Dziennus, M., Graindorge, T., Ortega, E., Graindorge, M., Merat, N.: Acceptance of Automated Road Transport Systems: An Adaptation of the UTAUT Model. Transportation Research Procedia 14, 2217-2226 (2017)
19. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
20. Lee, M.-C.: Factors Influencing the Adoption of Internet Banking: An Integration of TAM and TPB with Perceived Risk and Perceived Benefit. Electronic Commerce Research and Applications 8, 130-141 (2009)
21. Davis, F.D.: A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Massachusetts Institute of Technology, Doctoral Dissertation, 1-291 (1985)
22. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982-1003 (1989)
23. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MISQ 27(3), 425-478 (2003)
24. Taylor, S., Todd, P.A.: Understanding Information Technology Usage: A Test of Competing Models. ISR 6(2), 144-176 (1995)
25. Cheng, T.C.E., Lam, D.Y.C., Yeung, A.C.L.: Adoption of Internet Banking: An Empirical Study in Hong Kong. Decision Support Systems 42(3), 1558-1572 (2006)
26. Giang, P.T., Trang, P.T., Yen, V.T.: An Examination of Factors Influencing the Intention to Adopt Ride-Sharing Applications: A Case Study in Vietnam. Imperial Journal of Interdisciplinary Research 3(10), 618-623 (2017)
27. Agarwal, R., Prasad, J.: Are Individual Differences Germane to the Acceptance of New Information Technologies?. Decision Sciences 30(2), 361-391 (1999)
28. Wang, Y.S., Lin, H.H., Luarn, P.: Predicting Consumer Intention to Use Mobile Service. ISJ 16(2), 157-179 (2006)
29. Chowdhury, S., Ceder, A.A.: Users' Willingness to Ride an Integrated Public-Transport Service: A Literature Review. Transport Policy 48, 183-195 (2016)
30. Ohta, H., Fujii, S., Nishimura, Y., Kozuka, M.: Analysis of the Acceptance of Carsharing and Eco-Cars in Japan. International Journal of Sustainable Transportation 7(6), 449-467 (2013)
31. Dütschke, E., Peters, A.: Why are Individuals Likely to Change to Sustainable Modes of Transport like Carsharing and Electric Vehicles? An Empirical Analysis. Working Paper: Fraunhofer Institute for Systems and Innovation Research (2017)
32. Fleury, S., Tom, A., Jamet, E., Colas-Maheux, E.: What Drives Corporate Carsharing Acceptance? A French Case Study. Transportation Research Part F 45, 218-227 (2017)
33. Efthymiou, D., Antoniou, C., Waddell, P.: Factors Affecting the Adoption of Vehicle Sharing Systems by Young Drivers. Transport Policy 29, 64-73 (2013)
34. Cheng, X., Zhu, R., Fu, S.: Modeling the Motivation of Users' Sharing Option: A Case Study Based on A Car-Sharing Digital Platform. Wuhan International Conference on e-Business, pp. 1-9 (2016)

35. Wang, Y., Wang, S., Wang, J., Wei, J., Wang, C.: An Empirical Study of Consumers' Intention to Use Ride-Sharing Services: Using an Extended Technology Acceptance Model. Transportation *in press* (2018)
36. Delhomme, P., Gheorghiu, A.: Comparing French Carpoolers and Non-Carpoolers: Which Factors Contribute the Most to Carpooling?. Transportation Research Part D 42, 1-15 (2016)
37. Wright, S., Cellina, F., Bulgheroni, M, Cartolano, F., Lucietti, L, van Egmond, P., van Wijngaarden, L.: Public Acceptance of SocialCar, a New Mobility Platform Integrating Public Transport and Car-Pooling Services: Insights from a Survey in Five European Cities. In: Proceedings of the 7th Transport Research Arena 2018, pp. 1-10 (2018)
38. De Souza Silva, L.A., de Andrade, M.O., Maia, M.L.A.: How Does the Ride-Hailing Systems Demand Affect Individual Transport Regulation?. Research in Transportation Economics *in press* (2018)
39. Tan, F.T.C., Tan, B., Lu, A., Land, L.: Delivering Disruption in an Emergent Access Economy: A Case Study of an E-Hailing Platform. CAIS 41(1), 1-21 (2017)
40. Ruangkanjanases, A., Techapoolphol, C.: Adoption of E-hailing Applications: A Comparative Study between Female and Male Users in Thailand. Journal of Telecommunication, Electronic and Computer Engineering 10, 1-10 (2018)
41. Lee, Y., Kozar, K.A., Larsen, K.R.: The Technology Acceptance Model: Past, Present, and Future. CAIS 12(1), 752-780 (2003)
42. Turner, M., Kitchenham, B., Brereton, P., Charters, S., Budgen, D.: Does the Technology Acceptance Model Predict Actual Use? A Systematic Literature Review. Information and Software Technology 52(5), 463-479 (2010)
43. Wixom, B.H., Todd, P.A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. ISJ 16(1), 85-102 (2005)
44. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2010)
45. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. ISJ 2(3), 192-222 (1991)
46. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in Online Shopping: An Integrated Model. MISQ 27(1), 51-90 (2003)
47. Doll, W.J., Hendrickson, A., Deng, X.: Using Davis's Perceived Usefulness and Ease-of-Use Instruments for Decision Making: A Confirmatory and Multigroup Invariance Analysis. Decision Sciences 29(4), 839-869 (1994)
48. Peterson, R.A.: A Meta-Analysis of Cronbach's Coefficient Alpha. Journal of Consumer Research 21(2), 381-391 (1994)
49. Chiu, C.M., Wang, E.T.G.: Understanding Web-based Learning Continuance Intention: The Role of Subjective Task Value. Information & Management 45, 194-201 (2008)
50. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 18, 39-50 (1981)
51. Jahangir, N., Begum, N.: The Role of Perceived Usefulness, Perceived Ease of Use, Security and Privacy and Customer Attitude to Engender Customer Adaptation in the Context of Electronic Banking. African Journal of Business Management 2(1), 32-40 (2008)
52. Subramanian, G.H.: A Replication of Perceived Usefulness and Perceived Ease of Use Measurement. Decision Sciences 25(5-6), 863-874 (1994)
53. Agatz, N., Erera, A., Savelsbergh, M., Wang, X.: Optimization for Dynamic Ride-Sharing: A Review. EJOR 223(2), 295-303 (2012)
54. Weigel, F.K., Hazen, B.T., Cegielski, C.G., Hall, D.J.: Diffusion of Innovations and the Theory of Planned Behavior in Information Systems Research: A Metaanalysis. CAIS 34, 619-636 (2014)

Modeling Delay Propagation and Transmission in Railway Networks

David Rößler[1], Julian Reisch[2] and Natalia Kliewer[1]

[1] Freie Universität Berlin, Department of Information Systems, Berlin, Germany
{david.roessler,natalia.kliewer}@fu-berlin.de
[2] DB Netz AG, Department for Timetabling and Capacity Management, Berlin, Germany
julian.reisch@deutschebahn.com

Abstract: In railway scheduling, the planning of time supplements is crucial to the robustness of the resulting timetable. Time supplements as a means to accommodate for train delays are often distributed according to operation rules and based on experience. A part of the project for strategic schedule optimization at *DB Netze* aims at improving the supplements distribution through learning of structures of delay propagation and transmission from historical railway operation data. The work at hand focuses on delay transmissions between trains. It employs correlations and correlation network analysis to identify and analyze these knock-on delays and to develop logical precedence orders of trains at certain operation points which can in turn be used in a sequential calculation of single train delay propagation. Furthermore, it endeavors to establish a basis to identify strongly connected groups of trains and stations, thus forming relevant subnets for further analysis.

Keywords: data analytics, correlation network analysis, delay management, rail transportation, railway timetabling

1 Introduction

The German railway network is with a total of about 38,000 track kilometers the largest in Europe. The complexity of the timetabling process arises from three conflicting objectives that need to be balanced out. First, the capacity on the infrastructure, that is, allocation of trains to tracks, is to be maximized. Secondly, trains should operate with reasonably tight schedules and thirdly, the timetable must be robust against minor disruptions.

This paper contributes to a research project that aims at optimizing the third aspect, the robustness, meaning that the number of trains and also the traveling times are fixed. One way to cope with minor disruptions and hence improve robustness is to include slack in the timetable, that is time supplements. The trade-off between scheduling slack to achieve robustness against unforeseen events and the goal to realize a schedule as

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

efficient as possible and to operate as many trains as possible is evident. To improve upon this trade-off by identifying optimization potentials within the current distribution of supplements is the greater goal.

The project this work is part of is structured in the following way: First, we analyze delay propagation through the network of train operations. Clearly, if a train is delayed at a certain station, then this delay might still have an impact on the amount of delay in its next operation point. Moreover, delay may also be transmitted from one train to another. This happens for manifold reasons, e.g. passenger transits, track blockage and so forth. We implement a correlation network to detect the most important inter-train interdependencies. Then, in the second part of the project, we model delay propagation probabilistically, but only within one train operation and through the formerly detected interdependencies between trains. This gives us the possibility to see what happens, if the probabilities for a delay change, e.g. if trains always depart on time and so on. Furthermore, this model will be the basis of an optimization task, which will be the third and last part of our project. We will apply algorithms that modify the distribution of time supplements and check whether delays might propagate less, thus yielding a marginally increased punctuality.

The work at hand contributes to this project in the following way: We develop and deploy a novel approach to modeling railway train interdependencies with respect to the propagation and absorption of delays. As [1] have presented in their comprehensive survey, railway data analytics can benefit from employing *Big Data* methods. With a focus on scalability for the included development task, our paper explores and selects procedures and tests them on an exemplary large data set from German railway operations. This enables us to detect which trains have a dependency significant enough so that it should be included in the delay propagation model. Furthermore, the delay networks approach yields us the crucial trains and stations where a delay has a huge impact on many other trains in the network. It will prove wise to first optimize the punctuality of these trains, for instance in an initial solution of a future optimization algorithm.

The outline of this paper will be as follows. In Section 2 we introduce the data from railway operations and motivate our choice of data selection. We then give a brief insight in how we clean our data with respect to outliers, missing values and seasonality. What is next, we present our model for delay transmission. More precisely, we analyze the influence of the absolute value of delays of one train to the change of delay of a succeeding train. In Section 3, we start by discussing two measurements of this influence, namely the Pearson and the Kendall correlation coefficient. This gives rise to a delay transmission network that utilizes these measurements as weights on the edges of a network graph model. Section 4 ends with an analysis of a select example and the validation of the overall results. Finally, we give a short conclusion in Section 5.

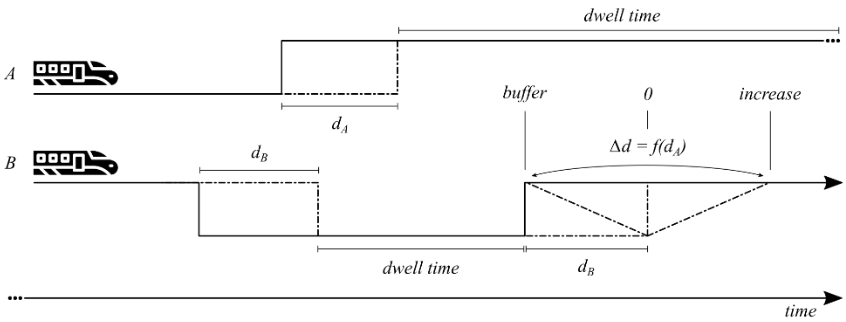


Figure 1. Model for delay transmission between two trains. Train A arrives with a delay d_A and train B with d_B. The change in delay upon departure of train B is ranged between *buffer*, where $\Delta d = -d_B$, no change, where $\Delta d = 0$, and *increase*, in which case Δd.
Source: own compilation.

2 Methods

We focus on the effects of the total delay of an arriving train on the ability of any departing train to reduce prior delay by using up time buffer or the necessity to build-up additional delay. Figure 1 illustrates this idea. We denote the *total delay upon arrival* as d and the *change in delay* as Δd. The suspected relationship can be expressed by $\Delta d_{Dep} = f(d_{Arr}) := \Theta \cdot d_{Arr}$, where Θ is an arbitrary function which yields a measure of association.

2.1 Data Selection: Region and Traffic Type

As an example, for our analysis, we choose the long-distance train network in the south western region of Germany where we deal with two major railway corridors. The first one is from Basel to Frankfurt and the other one from Stuttgart to Cologne. These corridors meet in Mannheim, where a transfer is possible as the different long-distance train lines are synchronized there. This synchronization is the first reason to choose this region as we expect interdependencies of the long-distance trains there. Furthermore, the two corridors are highly frequented so that the capacities are fully saturated, and an optimization of slack is helpful. We discretize the corridors by flag stops and omit signals which do not involve a regular stop. Accordingly, this study focuses on the southern part of the first corridor with stations *Offenburg (RO)*, *Baden-Baden (RBB)*, *Karlsruhe (RK)*, and *Mannheim (RM)*, the latter two being the cities with the second and third largest population in Baden-Württemberg, whereas the former stations are stereotypical for smaller sized cities. A large variety of train types is moving along the German railway network. We generally distinguish between regular and irregular and long- and short-range passenger trains and freight trains. Since this work's approach is to analyze regular train encounters with a focus on temporal precedence, and since freight traffic behaves structurally irregular, this work will abstract from freight trains and solely consider regularly running passenger trains - both long and short distance.

2.2 Data Cleaning: Outliers, Missing Data, STL

Before the actual data analysis can be performed, data must be cleaned and, if necessary, transformed[1]. The data preparation procedure for the work at hand comprises of the following key activities: *outlier detection and removal*, *handling of missing data* and, finally, *correction for trend and seasonality*.

Outliers. The univariate distribution of the total delay of a train and the change in delay cannot generally be expected to be symmetric. Positive delays occur more frequently than negative ones. The latter would imply that a train runs faster than necessary to maintain punctuality, which is inefficient and generally unwanted, whereas positive delay is an unwanted though inevitable phenomenon. Analyses of train delays of passenger trains in the Netherlands [2, 3] showed right-skewed distributions of delay across trains. In fact, delays for railway trains are commonly modeled as log-normal, exponential, Weibull, or gamma distributed random variables [4]. In our data, the same asymmetry can be observed (see Table 1). Accordingly, the empirical distributions of our random variables are skewed, and their location parameters and moments shifted unevenly. Hence, symmetrical outlier detection methods, like Tukey's fences [5], will falsely classify extreme observations on the heavy tailed side of the distribution as outliers and fail to detect outliers on the steep side. For this, [6] used the *medcouple* (MC) measure to adjust Tukey's fences for skewness. The medcouple, as proposed in [7], is a robust and efficient measure for skewness with a contamination breakdown barrier as high as 25%[2]. It can be calculated in $O(n \log n)$ time. Thus, MC is a best compromise between robustness, complexity and skewness detection performance. For un-skewed distributions, both the basic and adjusted method yield the same results. However, if the data distribution is skewed, then the adjusted boxplot method accounts for the skewness, even in the case of contamination due to the presence of far outliers.

	$[S > 0.05]$	$[-0.05 \leq S \leq 0.05]$	$[S < -0.05]$
Total delay	3104	2965	247
Change in delay	1969	3328	1056

Table 1. Number of trains, which are either right-skewed, symmetrical, or left-skewed. The skewness S was measured by means of the medcouple, showing that the majority of trains show exhibit right-skewness or approximate symmetry for both d and Δd.

Using this method *13.95%* of all cases were marked as outliers caused by either of our interesting variables. Out of all observations for the delay upon arrival *5.87%* and for

[1] Technologies like *bootstrapping* exist, where these preparation steps are not necessary, however, we expect them not to scale well.

[2] In 8 the asymptotic breakdown point of an estimator T is derived as $\epsilon^* = \epsilon^*(F_0, T) = inf\{\epsilon \mid b(\epsilon; X, T) = \infty\}$, where ε is the amount of deviant (i.e. contaminated) data, b the bias function, and ε* the minimum level of for which the estimate bias becomes infinite.

change in delay *8.34%* where marked as outliers both of which lie well within the breakdown range of MC.

Missing Data. Based on expert knowledge, many occurrences of missing observations in the dataset have been rectified prior to the import into the *RDBMS*[3]. However, missing values for both variables persist. In order to decide upon the correct means of dealing with missing values, first, the cause of their absence must be identified. In data mining, three mechanisms of missing data are distinguished: *missing at random* (MAR), *missing completely at random* (MCAR), *missing not at random* (MNAR) [9]. If the occurrence of missing values is completely unrelated to the manifestations of the variable itself and other observed data, the underlying mechanism is MCAR and can be considered as *ignorable* [10]. The occurrence of missing values in the data seems unrelated to other relevant variables in the data set and is due to the lack of contradicting evidence assumed to be caused by MCAR. MCAR with as low occurrence rates (*3.26%* of the values for *delay upon arrival* and *4.95%* of the values for *change in delay*) as in the present data can be treated with the *complete-case* method, by which only complete observations will be regarded in the analysis[4].

Trend & Seasonality. We can identify systematic differences in delays and the delay

Total delay	$\overline{x}$	2^{nd} *Quart.*	$\widetilde{x}$	3^{rd} *Quart.*
Mon	75.75	-6.00	47.00	173.00
Tue	104.43	4.00	69.00	207.00
Wed	129.21	9.00	77.00	218.00
Thu	160.43	9.00	77.00	213.00
Fri	161.18	10.00	81.00	223.00
Sat	153.36	10.00	83.00	226.00
Sun	92.92	-4.00	52.00	176.00
Change in delay	$\overline{x}$	2^{nd} *Quart.*	$\widetilde{x}$	3^{rd} *Quart.*
Mon	2.78	-27.00	-1.00	17.00
Tue	-0.45	-32.00	-3.00	15.00
Wed	-4.03	-34.00	-4.00	14.00
Thu	-4.50	-34.00	-4.00	14.00
Fri	-2.31	-33.00	-4.00	15.00
Sat	-3.44	-33.00	-4.00	15.00
Sun	-0.31	-28.00	-1.00	16.00

Table 2. Average total delay and change in delay by days of week.

differences between days of the week. On average, the total delay upon arrival at an operation point peaks on Fridays. Analogously, the ability to reduce delay seems to

[3] Relational Database Management System: *Maria DB* (https://mariadb.org/).

[4] We have added binary dummies, representing missingness in the total delay and in the change in delay, and then checked, whether they are correlated with each other and with the original values. Furthermore, we have performed a pairwise contingency test for the dummies and cardinal variables (train type, source station), and only found evidence for a, though significant, rather weak association between them (*Cramér's v* $\leq .30$). We found that missingness for the total delay and change in delay are strongly connected ($\rho = .782$). In most occurrences of missingness in either variable the other was missing as well, which supports the use of the complete case method. At the same time, missingness appears to be virtually unrelated to the absolute values of the delay variables.

decrease (see Table 2). The Wilcoxon rank sum test to compare medians of unpaired samples shows that the daily means are significantly shifted[5]. In addition, temporary rescheduling might lead to a-cyclical local and global trends, which must be addressed. To model the described trends and weekly patterns, we use an additive component model. The decomposition is achieved using a combined approach called *seasonal-trend decomposition procedure based on LOESS* (STL), as presented in [11].

2.3 Data Engineering: Cumulated Delay, Train Encounters

As the change in delay of a train after its departure shall be analyzed, data rows which constitute a departure process are merged in order to obtain the cumulated *change in delay*. Furthermore, as this study focuses on associations between train delays at specified operation points and at specific times, encounters of one train with another are obtained by matching the time of arrival of an incoming train with the time of departure the departing train at an operation point. Lastly, relevant and valid subsets must be selected with respect to the size of the resulting sub-samples.

Cumulative Change in Delay. The process of a train arriving and departing at an operation point is established as a sequence of events connected by activities called an *activity-event network* ([12, 13], similarly). A simplified version of such a network is depicted in Figure 2. The nodes in the graph represent signal passing events and links between event nodes indicate the transitions from one activity to the other, e.g. the train moving from signal to signal. Each transition can result in a change in delay Δd. The total delay d of a train i upon arrival can be represented as the difference between its scheduled time of arrival $\hat{t}^{a}_{i}$ and its actual time of arrival t^{a}_{i}, at that time: $d_i(\hat{t}^{a}_{i}) = t^{a}_{i} - \hat{t}^{a}_{i}$ This approach is rather straightforward and the resulting delay indiscriminately encompasses primary and secondary delays, which the train in question has accumulated during its course up until arriving at the respective operation point. The change in delay during the train's departure follows a slightly more complex pattern. As the reader can see in Figure 2, a change in delay can be the result of transitions (2) or (3) simultaneously. The delay change during one transition from t_k to t_j, $\Delta d_i(t_k,t_j)$, can be calculated as $\Delta d_i(t_k,t_j) = d(t_k) - d(t_j)$. For the work at hand, it is considered negligible which transition causes the build-up or decrease in delay, given that it is induced by the delayed arrival of another train. We accept the possibility for causes of the change in delay (e.g. passengers boarding and alighting), however, abstract from it. Because there is a systematic element in how the change in delay is distributed across phases, only the combination of all phases is considered. For simplicity, in this work a linear combination, i.e. the sum, will be applied: $\Delta d_i(t_k, t_{k-n}) = \sum_{j=1}^{n} \Delta d_i(t_k, t_{k-j})$, where $n \in \mathrm{N} \wedge n < k$ Putting the above equations together and simplifying yields the *cumulated change in delay*: $\Delta d_i(t_k,t_{k-n}) = d(t_k) - d(t_{k-n})$.

[5] As can be guessed from Table 2, the medians for Fridays and Saturdays as well as Wednesdays and Thursday are not significantly different.

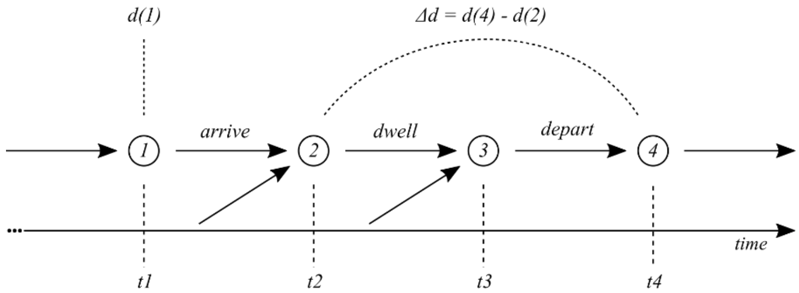


Figure 2. In (1) the train leaves the section prior to the station and enters its area. The train then arrives at the platform/station track (2) and leaves the platform in (3). In (4) the train leaves the station area for the next section. Between events (2) and (3) the train dwells at the platform for boarding and alighting or a conductor change etc. Source: own compilation based on [13].

Train Encounters. Extracting actual train encounters is an important preparatory step for the analysis. The following gives a concise definition of what is meant by the term *encounter*: For the purpose of this work, the arrival of one train *i* and the departure of a train *j* are considered an encounter, if the target operation point of *i* equals the start operation point of *j*, and, if the actual arrival time of i, t_i^a, was less than 10 minutes before the time of departure of j, t_j^a. Following this understanding, encounters imply temporal precedence, e.g. it is logically assumed that a train can only pass its delay on to trains, which depart at a later time, and never to trains that have already left the operation point. This assumption generally holds, however, the opposite direction is not inconceivable, as the dispatcher may always take action to create a situation in which one train passes its delay on to a preceding train and thus works as a time forwarding transmitter. In the further analysis, this inversion does not constitute an encounter.

Exclusion of Small Sub-Samples. For this paper, we consider only pairs of trains which have a minimum number of *30* encounters, using the method presented in [14] to determine exact sample sizes. These sizes depend on pre-estimated correlation coefficients, which are determined either through expert knowledge or prior research. For this work, the overall correlation between the d and Δd serves as an estimate for the expected magnitude of ρ and τ.

3 Constructing the Delay Transmission Network

3.1 Pearson's Product-Moment Correlation Coefficient ρ

For continuous variables which are at least interval-scaled *Pearson's product moment correlation coefficient* is a measure of choice. It gives the change in a random variable X which coincides with an increase or a decrease in another variable Y and vice-versa.

The coefficient is defined for the interval [−1,+1] and measures the strength and direction of a linear association between two variables. A negative sign indicates an anti-proportional relationship and a greater absolute value implies stronger association between the variables. If the coefficient is equal to 0, the two features do not exhibit any *linear* relationship. However, a non-linear function to describe their relationship might still exist [15]. The product-moment correlation coefficient for random variables X and Y is given by $\rho(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$ where x and y are realizations of the variables. Thus, the interpretation of the calculated coefficients is rather straightforward: Let $X = d^k_{ij}$ denote the absolute delay upon arrival of a train i at a station k in an encounter with a train j. Furthermore, let $Y = \Delta d^k_{ij}$ signify the change in delay of that train j in an encounter with the preceding train i. Then $\rho(X,Y)$ gives the amount of time by which Δd^k_{ij} would increase or decrease, if d^k_{ij} were to be increased or decreased by one second. In other words, ρ is the extent to which the delay of train i proportionally influences the aggregate delay build-up (or reduction) of train j during dwelling and/or departure. $\rho(X,Y) = 1$ means that an increase or a decrease by one second of delay in train i fully translates into an increase or a decrease in the change in delay by that same amount and $\rho(X,Y) = -1$ means the exact opposite.

3.2 Kendall's Rank Correlation Coefficient τ

For ordinal variables, *Kendall's rank correlation coefficient* is an appropriate measure. Its interpretation is very similar to Pearson's coefficient, however, it does not measure the linear relationship between two features, but whether both variables share the number of discordances, instead. We use a modified version of the coefficient, *tau-b*, which accounts for rank ties. The rank correlation coefficient τ_b, corrected for the bilateral presence of ties is defined as $\tau_b(X,Y) = \frac{S}{\sqrt{\left(\frac{n(n-1)}{2}-T_X\right)}\cdot\sqrt{\left(\frac{n(n-1)}{2}-T_Y\right)}}$ where T_X and T_Y denote the number of ties in the rank pairs of X and Y (in adaptation of the formula given in [16]). In this paper, Kendall's τ will serve as a verification instrument for Pearson's coefficient, as it is non-parametric and, other than ρ, is resilient even in the presence of far outliers [17].

3.3 Graph Theory & Network Analysis

This paper employs elements of graph theory and network analysis to further analyze relationships between trains. While a multitude of applications for network analysis in exploratory data analysis exists, if the network properties of the available data are evident [18], this toolbox has also found use in areas where graphical structures in the underlying data are less ostensible. Examples can be found in qualitative studies where covariates to an interesting outcome might exhibit multiple moderation effects [19], and, for some time now, in genetics where network-graphs are created based on correlations as a similarity measure for gene-expression states ([20, 21]). The foundation of a network structure is a graph G which comprises of *vertices* and *edges*. If two vertices are connected by an edge, they are called *adjacent*. Hence, G can be represented by its *adjacency matrix* $A_G = [a_{ij}]$ in which each a_{ij} represents the number of connections between two nodes i and j. Every graph H with vertices and edges from G restricted by the adjacency matrix A_G is a *subgraph* of G. A network is a graph-based structure and its interpretation specific to the application-domain – the *network graph*. It is formed from *nodes* (vertices) which are connected through *links* (edges) and extends the graph model with additional attributes for vertices and edges. A typical addition in the network context is the assignment of weights w_{ij} to each link. A *weight matrix* W_G would have the same shape as the adjacency matrix of the underlying graph. A network, in which the links have weights of arbitrary value, is called a *weighted* network. If links in the network connect ordered pairs of nodes exclusively, then the network (graph) is *directed* [22].

3.4 Constructing the Delay Transmission Network

The proposed network is based on a graph consisting of a set of vertices ($v_i \in V$) which each represent a respective train (number). Relationships between trains are expressed through a set of directed links $e_{ij} \in E\ \forall\ (v_i, v_j) \in (V, V)$ in the network graph. These directed links represent the dependency of the target node's change in delay from the source node's total delay. Weights in the constructed network are based on the correlation coefficients. In the constructed network graph, link weights are interpreted as similarities or the relative closeness between adjacent nodes. Hence, on the one hand, while negative weights are not generally inconceivable (for example [23]), they are implausible in the current application context. Negative values for the correlation coefficients, on the other hand, are very plausible, and must be dealt with prior to performing the network analysis. Otherwise, the resulting negative and positive weights might bias weighted and distance-based network measures. The application of soft-thresholding produces weights for the proposed network. In this work they are obtained based on the arithmetic means of the correlation coefficients γ: Let γ be the mean of the correlation coefficients ρ and τ and let exist an arbitrary number $\lambda \in <$ then the weight w_{ij} for the link connecting the nodes i and j is given by $w_{ij}^{p} = |\gamma(v_i, v_j)|^{\lambda}$. The resulting values fit the interval *[0,1[,* thus, preserving information on the strength of the respective correlation. Additionally, the exponent λ is included. Performing this operation from the Tukey-ladder of power-transformations [5] adds the ability to reduce tail weights in the coefficients' distributions.

3.5 Measuring Network Properties

In this paper, we use the node *strength* as an indicator for the importance of a node. The node strength respects the strength of ties with other nodes in the network and calculates as the sum of link weights [24]. Of particular interest for this paper is the out-strength $D_{in}(p_k) = \sum_{i \neq k} w_{ik}$ and in-strength $D_{out}(p_k) = \sum_{i \neq k} w_{ki}$ – analogous to the degree-measures.

An adequate means to identify possible moderation-effects, which certain trains might have, is betweenness centrality (C_B) as detailed in [25]. It can be considered a measures of a node's overall connectedness with the network[6]. The betweenness $b_{ij}(p_k)$ of a node k with respect to two other nodes i and j is defined as the ratio of the number of paths between i and j which contain k, and the number of all paths connecting i and j: $b_{ij}(p_k) = \frac{1}{g_{ij}} \times g_{ij}(p_k)$. *Betweenness Centrality* (BC) is the sum of a node's betweenness for all node pairs formed from the $n - 1$ other nodes in the network: $C_B(p_k) = \sum_{i=1}^{n} \sum_{j \in [1,i[} b_{ij}(p_k)$

4 Evaluation

In the following section, we take a look at some select examples and evaluate the approach. We have implemented it using *GNU-R*[7]. Furthermore, we have used *Gephi* and *Inkscape* for visualizations.

4.1 Results

In the subset around the four selected operation points, we can observe the following data as follows.

Number of Trains by Operation Point. Figure 2 shows the number of trains arriving

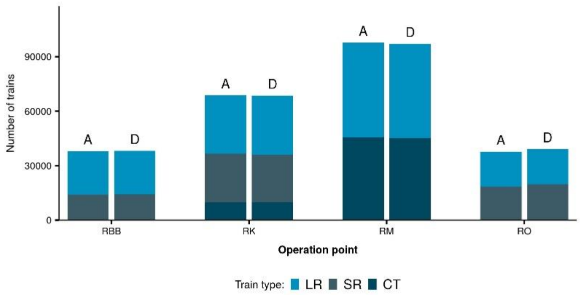


Figure 2. Number of trains arriving (A) at and departing (D) from the selected operation points. Source: own compilation.

and departing at the respective operation points and the distribution of different train types. As was expected, *Mannheim* and *Karlsruhe* handle much more than twice as much traffic as the two smaller operation points.

[6] The concept of betweenness was originally used as a measure of a person's ability to influence a group. A high betweenness would mean that a person is able to control the flow of information in the network.

[7] To name the central libraries: *reshape2*, *RMySQL*, *dplyr*, *broom* (data handling); *robustbase*, *stlplus* (outlier removal, detrending, de-seasonalization); *igraph* (network modeling); *ggplot2* (visualization).

Number of Encounters by Operation Point. Accordingly, we were able to extract encounters, as presented in Figure 3. The smaller operation points, *RO* and *RBB*, on the one hand, show a similar distribution having most trains being involved in well less than 1000 encounters. The larger operation points, *RM* and *RK*, on the other hand, appear to be very different. The *Mannheim* plot looks like a scaled version of that of the former two. At *RK*, most trains have even fewer encounters than in the smaller operation points, and at the same time, very few trains have a great number of encounters. This would appear to be a result of regular strong peaks in the number of arriving and departing trains.

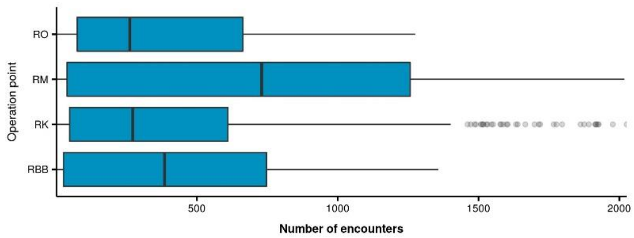


Figure 3. Number of encounters by train at the selected operation points. Given is the count of all encounters a train is involved in (as both arriving and departing). Source: own compilation.

Correlation Coefficients. The Kendall and Pearson correlation coefficients' distributions show similar patterns, with the median indicating a weak negative correlation. Both have a cluster in the weak to medium positive correlation range. In Figure 4, these train couples appear as outliers on the right side of the scale. However, Kendall's τ doesn't reach the extremes of the scale $(-1,1)$ as much as ρ does.

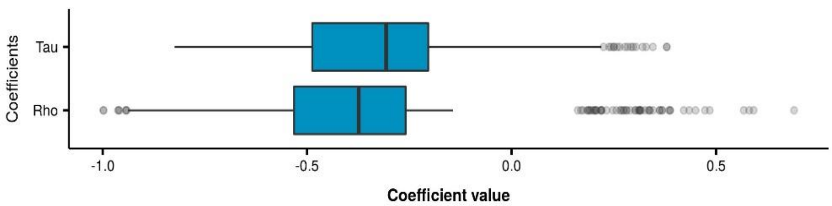
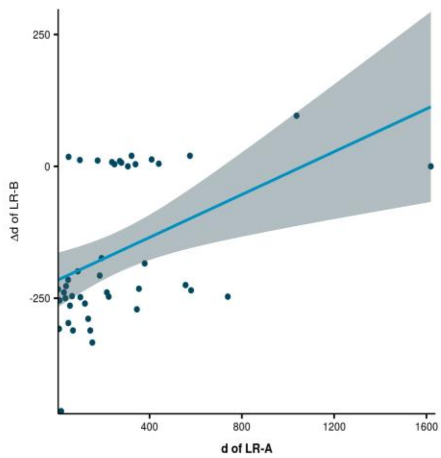


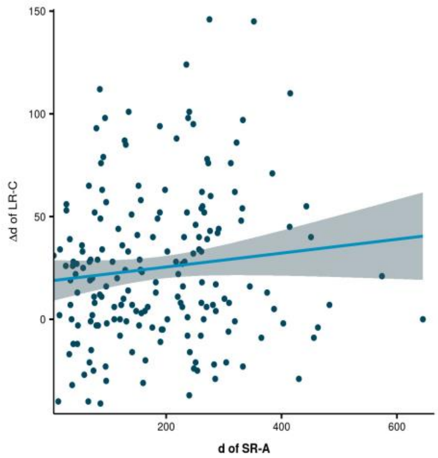
Figure 4. Comparison of the two correlation measures. Source: own compilation.

4.2 Examples

To test the approach, the described network was constructed for the *Mannheim* operation point. As for the link weights, we regard only significant correlation coefficients for τ (with $\alpha = 0.05$) and with maximum CI range of 0.3 for ρ. The resulting density distribution is almost Gaussian with a mean of -0.12. There are several strongly connected trains for which an increase in delay upon arrival in one train coincides with an increase in the change in delay in the other. In Figure 6, the relationship between different trains is presented. There clearly exists a positive relationship between delay



(a) LR-A on LR-B



(b) SR-A on LR-C

Figure 5. Scatter plot for encounters of trains with positive correlation in *Mannheim*. LR:= long-distance train; SR:= regional train. Source: own compilation.

upon arrival and the change in delay. Figure 5 (a) plots the total delay of long-distance passenger train $LR\text{-}A$ against the change in delay of another long-distance passenger train $LR\text{-}B$. Figure 5 (b) plots the delay of a regional train $SR\text{-}A$ against the delay change of a long-distance passenger train $LR\text{-}C$. Both situations seem plausible. The trains' respective planned arrival and departure in each coupling are at least 7 minutes apart, qualifying as a connection. While the correlation between the trains in pair (a) is relatively strong by comparison, that in (b) finds week support, as the dots are rather unevenly distributed. It is possible, nonetheless, that the correlation in (b) is due to a feeder train relationship, i.e. that $SR\text{-}A$ is a feeder for $LR\text{-}C$. Table 3 represents the network characteristics of our examples. The out-strength of all long-distance passenger trains is in the 3rd and 4th quartile (median: 0.32) of the strength distribution for $Mannheim$ and can be considered as medium to highly influential for this operation point. Most trains, which exhibit strong outgoing links, are long-distance passenger trains as well. $LR\text{-}A$ is also a highly influential train; however, it has incoming links. Yet, it is the only train with a link to $LR\text{-}B$, which is the target of only one incoming edge. This example is remarkable, as $LR\text{-}B$, on many days, has no change in delay, at all. On other days, it reduces delay, probably prompted by its own delays. However, its ability to decrease delay seems to be negatively related to the total delay of $LR\text{-}A$. Similarly, $LR\text{-}C$'s change in delay is correlated only with the total delay of $SR\text{-}A$. However, the latter exhibits a higher in-strength and little out-strength. In the respective figure, we can ascertain that the correlation is rather weak ($\rho = .123$ and $\tau = .195$), yet significant.

	Out-Strength	In-Strength	Out-Degree	In-Degree	Betweenness
LR-A	0.68	0.48	3.00	2.00	0.00
LR-B	0.51	0.30	2.00	1.00	0.00
SR-A	0.16	0.57	2.00	2.00	0.00
LR-C	0.65	0.16	2.00	1.00	0.00

Table 3. Exemplary trains and their network characteristics.

Betweenness values are generally very small. With means at *.012* and *.031* and maximum values at *.17* and *.005*, moderation effects for the delay transmission appear negligible. This seems to be due to the fact, that the network is not well connected. The observable formation of cliques indicates that the transmission of delays is restricted to certain "areas" of times during the day.

4.3 Validation

To validate the results, we have extracted a sub-sample from the train encounters by randomly selecting *70%* of all encounters of each pair of trains. This serves as the training set on which we perform the analysis, as described above. The remaining *30%* serve

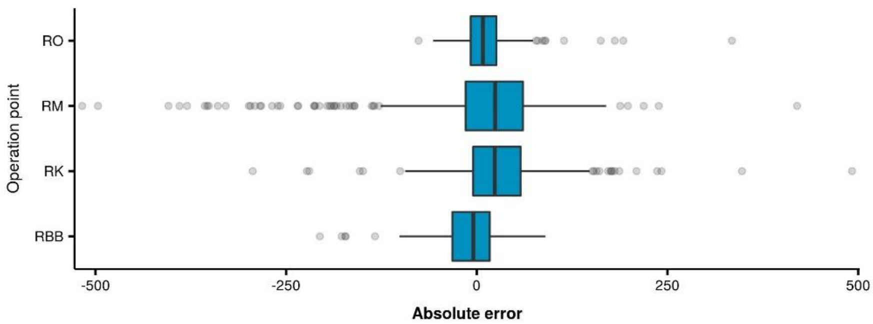


Figure 8. Boxplots for the distribution of differences for predicted vs. observed Δd for each pair of trains at the respective stop (10%-trimmed). Predictions were obtained by multiplying the total delay of the incoming trains with the estimated Pearson correlation coefficients.

as our test set. Figure 8 shows the distributions of differences at the four focal operation points. As can be seen, the larger operation points *Mannheim* and *Karlsruhe* exhibit slightly higher errors for all train pairs (i.e. lower accuracy) and higher variance in means than the smaller ones. However, the means are kurtotically centered around a gravity center close to zero. The overall *mean squared error* (MSE) for the predicted *Δd* is *149,999.59* (vs. *265,952.12* with mean values). Many of the individual models perform very poorly, however, some perform well - with the best R^2 of *70.70%.*

5 Discussion & Conclusions

As was stated, as of now, no distinction is made between secondary and primary delays. This is simply due to the fact, that recognizing true secondary delay build up, would go beyond the scope of the research task completed in this work. Discerning primary and secondary delays requires additional modeling, an approach that is discussed further in this section. Furthermore, our understanding of negative correlation coefficients is somewhat inconclusive. A working hypothesis is that negative correlations indicate discontinuities in the development of the change in delay, such as train order swaps; such that up to a certain arrival delay, transmission occurs until a threshold is reach, when transmission decreased towards *0*. The correlation coefficient might then be negative. In addition, further sophistication in the data preparation process or just broadening the data selection might consolidate interpretability and facilitate a better understanding. In the work at hand, an approach for the analysis of train delay propagation was demonstrated. As a result, train interactions can be determined for selected railway networks. These will be used as inputs at subsequent project stages, where we plan to use these inputs to retrieve a computation order for a by-train-optimization of time supplements. Further validation and verification will be part of further project stages. This involves expert evaluations and the inclusion of information on passenger movements, which has not been available to us, so far.

6 Acknowledgements

We are thankful for the historical operational data kindly provided by *DB Netze* and the support we have received.

References

1. Ghofrani, F., He, Q., Goverde, R.M.P., Liu, X.: Recent applications of big data analytics in railway transportation systems: A survey. Transportation Research Part C: Emerging Technologies 90, 226–246 (2018)
2. Goverde, R.M.P.: Punctuality of railway operations and timetable stability analysis. Netherlands TRAIL Research School (2005)
3. Yuan, J.: Stochastic modelling of train delays and delay propagation in stations. Eburon Uitgeverij BV (2006)
4. Schranil, S.: Prognose der Dauer von Störungen des Bahnbetriebs. ETH Zurich (2013)
5. Tukey, J.W.: Exploratory data analysis. Reading, Mass (1977)
6. Hubert, M., Vandervieren, E.: An adjusted boxplot for skewed distributions. Computational statistics & data analysis 52, 5186–5201 (2008)
7. Brys, G., Hubert, M., Struyf, A.: A robust measure of skewness. Journal of Computational and Graphical Statistics 13, 996–1017 (2004)
8. Huber, P.J., Ronchetti, E.M.: Robust statistics. John Wiley & Sons, Inc, New York (2009)
9. Little, R.J.A.: A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association 83, 1198–1202 (1988)
10. Rubin, D.B.: Inference and missing data. Biometrika 63, 581–592 (1976)

11. Cleveland, R.B., Cleveland, W.S., Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6, 3 (1990)
12. Schöbel, A.: Integer programming approaches for solving the delay management problem. In: Algorithmic methods for railway optimization, pp. 145–170. Springer (2007)
13. Büker, T., Seybold, B.: Stochastic modelling of delay propagation in large networks. Journal of Rail Transport Planning & Management 2, 34–50 (2012)
14. Bonett, D.G., Wright, T.A.: Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika 65, 23–28 (2000)
15. Taylor, R.: Interpretation of the correlation coefficient: a basic review. Journal of diagnostic medical sonography 6, 35–39 (1990)
16. Szmidt, E., Kacprzyk, J.: The Spearman and Kendall rank correlation coefficients between intuitionistic fuzzy sets. In: EUSFLAT Conf, pp. 521–528 (2011)
17. Long, J.D., Cliff, N.: Confidence intervals for Kendall's tau. British Journal of Mathematical and Statistical Psychology 50, 31–41 (1997)
18. Butts, C.T.: Revisiting the foundations of network analysis. science 325, 414–416 (2009)
19. Voth-Gaeddert, L., Cornell, D.: Improving health information systems in Guatemala using weighted correlation network analysis. In: Global Humanitarian Technology Conference (GHTC), 2016, pp. 686–693 (2016)
20. Ruan, J., Dean, A.K., Zhang, W.: A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC systems biology 4, 8 (2010)
21. Zhang, B., Horvath, S.: A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology 4, 1–45 (2005)
22. Bondy, J.A., Murty, U.S.R., others: Graph theory with applications. Citeseer (1976)
23. Schwarz, A.J., McGonigle, J.: Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. Neuroimage 55, 1132–1146 (2011)
24. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social networks 32, 245–251 (2010)
25. Freeman, L.C.: Centrality in social networks conceptual clarification. Social networks 1, 215–239 (1978)

What is the impact of company specific adjustments on the acceptance and diffusion of logistic standards?

Daniel Neuß[1], Samuel Pfister[1], Adrian Hofmann[1], and Axel Winkelmann[1]

[1] Julius-Maximilians-Universitaet, Chair of Business Management and Business Information Systems, Wuerzburg, Germany
{daniel.neuss,samuel.pfister,adrian.hofmann,axel.winkelmann}@uni-wuerzburg.de

Abstract. Transportation Networks and Logistics 4.0 work on the basis of integrated systems. These systems are enabled by standards which are widely used for the communication between IT-Systems. Due to individual requirements, standards are customized by companies. The extent of individualization has not yet been investigated. Therefore, we applied an empirical analysis on two mature logistic communication standards to identify the standard-application gap by calculating the interoperability between the applied standards. Within these results, we expand the standardization theory by simulating the network effects and the standard diffusion and quantify the impact of the company specific individualization of logistic standards. The identified findings help to improve the standard diffusion model and to improve the development of logistic standards.

Keywords: network effects, standardization, standard deviation, diffusion, logistic standards

1 Introduction

Global supply chains work with integrated logistics systems [1]. This integration offers several performance benefits [2], and is necessary for concepts like Industry 4.0 [3]. To accomplish a certain level of integration, standardization is necessary [4], [5]. Standards therefore have economic advantages like positive network externalities [6], [7] economics of scale [8] and the reduction of transaction costs [9], which are widely discussed theories. Theoretical research on standards deals to a large extent with network effects. Participants benefit from the advantages if the number of participants increases [10], [11]. Specifically, this means that the benefit for a standard user grows stronger if more actors implement the standard. This is combined in the standardization model from Weitzel et al. where the adoption decision is taken as soon as the advantages exceed the costs [11]. Consequently, the decision to introduce a standard is strongly linked to the expectation of how the other actors in the corresponding environment (e.g. within a certain industry) decide, which is known as the standardization problem. In conclusion, the adoption of a standard is more influenced by the anticipated diffusion than on

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

its actual direct benefit, i.e. its quality [11]. In the theoretical concepts of standardization, deviations have not yet been taken into account. This is based on the assumption that a standard is either not or fully implemented, and then becomes perfectly interoperable (and thus compatible). This simplifying assumption leads to a significant theory-practice gap. Therefore, we want to address the hypothesis that there is a direct impact on the standardization deviation, e.g. network effects, due to an interoperability gap.

In this article, we want to close this gap by extending the standardization model developed by Weitzel et al. Therefore, we introduce the interoperability metric from Zhu and Whu [12] and apply it with real data to derive the real interoperability of EANCOM D.96A standard messages. This mature and widely used standard defines several communication massages which are used in different parts of the supply chain. In the next step, we simulate the impact of this metric on the diffusion of the standards and network effects to conclude with an outlook of our future research.

2 Implications of company specific standard adjustments on network effects

Interoperability describes the agreement which exists if, for example, two business partners use the same standard. In practice, this results in different variants, which means that only a certain part of the defined data elements is used and, if necessary, individually defined elements are added. The incompatibilities are eliminated by bilateral coordination within the implementation of the standard. Interoperability measures the extent to which the data elements of two instances overlap (1) [12]. It describes the interoperability values I_{ij} between two actors i and j, where D_i is the set of data elements required by user i. The interoperability values in this paper are calculated from approximately fifty thousand messages of EANCOM standards and result in the arithmetic mean for orders 0.7842 (ORDERS D.96A) and shipping notifications (DESADV D.96A) of 0.8684. The messages were provided and anonymized by an IT service provider. To investigate the impact of the interoperability on to network effects and therefore on the further adaptation of standards, we use the network model by Weitzel et al. [11] and extend it by including the loss of efficiency produced by a lower interoperability.

$$I_{ij} = \frac{|D_i \cap D_j|}{\sqrt{|D_i||D_j|}} \tag{1}$$

The net gain of the whole network can be calculated as coordination efficiency (CE) which is the sum of the *ex post* network benefits of each actor E_i in a network with n actors, where c_{ij} is the direct network effect if the standard is used between actor i and j when the standard is fully interoperable (2). I_{ij} describes the interoperability of the standard between the actors, x_j is a binary variable that indicates whether j uses the standard and K_i represents the cost for an actor i to introduce the standard.

Table 1. Parameters used for the simulation

	Standard	*Model*	n	$\mu(c)$	$\sigma(c)$	$\mu(K)$	$\sigma(K)$	$\mu(I)$	$\sigma(I)$	$\tilde{I}$
S1	*(ideal)*	Base model	20	1000	200	variable	1000	1	0	1
S2	DESADV	Without I-Anticipation	20	1000	200	variable	1000	0.8684	0.0608	1
S3	DESADV	With I-Anticipation	20	1000	200	variable	1000	0.8684	0.0608	0.8684
S4	ORDERS	Without I-Anticipation	20	1000	200	variable	1000	0.7842	0.0866	1
S5	ORDERS	With I-Anticipation	20	1000	200	variable	1000	0.7842	0.0866	0.7842

$$CE = \sum_{i=1}^{n} E_i \qquad \text{with} \qquad E_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} c_{ij} I_{ij}^2 x_j - K_i \tag{2}$$

To decide if an actor adopts the standard (and therefore the value of x_j), an *ex-ante* calculation of the benefit can be used by each actor. The benefit has to be estimated since the actors don't know in advance who will adopt the standard (3). Here p_{ij} is the probability that actor i estimates for actor j to adapt the standard. ϕ_j is the number of j' s communication partners. In this first simulation we assume a full-density network, hence $\phi_i = n - 1$. To further evaluate the effects of interoperability we examine the effects of the actors knowing of interoperability losses (I-anticipation). In this case $\tilde{I} = I$, otherwise $\tilde{I} = 1$.

$$\tilde{E}_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} c_{ij} \tilde{I}_{ij}^2 p_{ij} - K_i \qquad \text{with} \qquad p_{ij} = \frac{c_{ji} \tilde{I}_{ji}^2 \phi_j - K_j}{c_{ji} \phi_j} \tag{3}$$

If the expected benefit $\tilde{E}_i > 0$, then actor i will implement the standard and $x_i = 1$. The results of the simulation are shown in Figure 1. Due to the possibility of an actor falsely estimating network effects that exceed the standardization cost, it is possible that the standard is adopted by not enough actors to fully utilize network effects, which leads to a negative net gain of the network. This is referred to as the *standardization gap.* It can be seen, that a lower interoperability impacts the network effects. If the cost for the introduction of the standard is low enough, there is no difference whether I-Anticipation is used or not. It should be noted that, even though standard adoption is higher at lower costs when the actors do not anticipate the interoperability, the standardization gap is more severe without I-Anticipation.

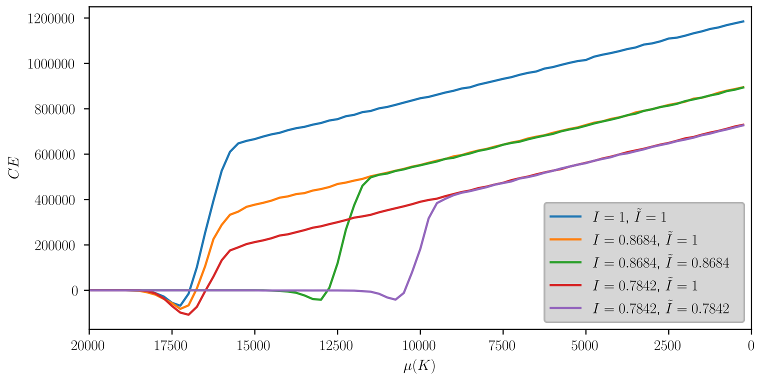


Figure 1. Simulated network effects

3 Conclusion and future work

Within the first simulation we could prove the effect of the interoperability on the standardization gap. In addition, we simulated the scenario with *I*-Anticipation and without. Therefore, we can conclude three major findings.

First, we can prove the hypothesis, that the individual and collectively achieved net benefit effects from a standard introduction with decreasing interoperability are constantly being reduced. Deviations in standards can lead to more or less severe economic losses. The **second** finding is that with medium to low standardization costs, it can provide economic sense to keep the actors unaware of possible interoperability losses. The area in which such a disinformation strategy offers a higher networkwide benefit increases as the expected interoperability decreases. However, in the case of medium-high cost, it is more advantageous for the actors to anticipate interoperability and therefore decide against a standard introduction. In the case of very low or high standardization cost, both strategies achieve identical results (i.e. no or complete standardization). The standardization gap can be divided into a centralization gap that has already been extensively researched and an interoperability gap that is shown for the first time in this paper and depends on the level of anticipated interoperability. This additional gap delays the diffusion of a standard, since the actors only consider an introduction if the cost of standardization is significantly lower, which concludes the **third finding.**

In conclusion there are two main points for further research. First, we want to expand the simulation to different standards to gather more practical insights on the deviation and use them for the simulation. In addition, we want to focus on the company and industry specific standardization gap. In the second step, we want to take the model extensions based on Weitzel et al. into account. These include, for example, the parallel availability of several standards with different interoperability factors, diffusion paths, different network topologies ($\phi_i \leq \text{n} - 1$) or several periods of time in which actors can subsequently make their individual decisions. Nevertheless, it is already clear from this study that losses in interoperability have a lot of theoretical implications for the theories of standardization, standard diffusion and network effects.

References

1. Liu, C.L., Lee, M.Y.: Integration, supply chain resilience, and service performance in third-party logistics providers. International Journal of Logistics Management. 29, 5–21 (2018)
2. Stank, T.P., Keller, S.B., Closs, D.J.: Performance benefits of supply chain logistical integration. Transportation Journal. 41, 32–46 (2001)
3. Lasi, H., Fettke, P., Kemper, H., Feld, T., Hoffmann, M.: Industry 4.0. Business & Information Systems Engineering. 6, 239–242 (2014)
4. Wook Kim, S.: Effects of supply chain management practices, integration and competition capability on performance. Supply Chain Management: An International Journal. 11, 241–248 (2006)
5. Goetschalckx, M., Vidal, C.J., Dogan, K.: Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. European Journal of Operational Research. 143, 1–18 (2002)
6. Farrell, J., Saloner, G.: Converters, compatibility, and the control of interfaces. The Journal of Industrial Economics. 40, 9–35 (1992)
7. Katz, M.L., Shapiro, C.: Network externalities, competition, and compatibility. The American Economic Review. 75, 424–440 (1985)
8. Dixit, A.K., Stiglitz, J.E.: Monopolistic competition and optimum product diversity. The American Economic Review. 67, 297–308 (1977)
9. Hudson, J., Jones, P.: Measuring the efficiency of stochastic signals of product quality. Information Economics and Policy. 13, 35–49 (2001)
10. Liebowitz, S.J., Margolis, S.E.: Are network externalities a new source of market failure? Research in Law and Economics. 17, 1–22 (1995)
11. Weitzel, T., Beimborn, D., König, W.: A unified economic model of standard diffusion: The impact of standardization cost, network effects, and network topology. MIS Quarterly. 30, 489–514 (2006)
12. Zhu, H., Wu, H.: Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. Decision Support Systems. 59, 351–360 (2014)

Robust Route Planning in Intermodal Urban Traffic

Matthias Ruß, Gunther Gust, and Dirk Neumann

University of Freiburg, Information Systems Research, Freiburg, Germany
{matthias.russ,gunther.gust,dirk.neumann}@ is.uni-freiburg.de

Abstract. Passengers value reliable travel times but are often faced with delays in intermodal urban traffic. To improve their mobility experience, we propose a robust route planning tool that provides routes guaranteeing a certain probability of on-time arrival and satisfying additional constraints. The constraints can limit the number of transfers, time-dependent trip costs and other relevant resources. To find such routes, we extend the time-dependent reliable shortest path problem by adding constraints on time-dependent and stochastic edge weights. An exact solution method based on multi-objective A* search is proposed to solve this problem. By applying our algorithm to a showcase featuring an actual city, we hope to answer relevant questions for policy-makers and contribute to smarter mobility in the future.

Keywords: Urban Mobility, Travel Time Reliability, Constrained Shortest Path Problem, Multi-Objective A* Search

1 Introduction

Urban transport passengers value reliable travel times [1]. However, travel times in urban transport are uncertain due to a range of factors such as the traffic situation or infrastructure defects. A robust route planning tool that takes reliability into account can therefore help to improve passengers' mobility experience in cities and enable smarter mobility. However, current tools typically don't consider reliability and base their advice on deterministic public transport schedules and expected car travel times.

Our contribution is the development of a route planner that finds robust routes in intermodal urban traffic. We define a network model that reflects intermodal specifics such as uncertain travel times that vary over time and a flexible set of constraints. These constraints can limit the number of transfers, time-dependent trip costs and other resources. Subsequently, we extend the work of Chen et al. [2] and develop an exact algorithm based on multi-objective A* search [3] that finds robust routes satisfying the constraints. Reliability is reflected by setting a minimum probability for on-time arrival.

2 Related Work

Reliable route planning is a stochastic shortest path problem [4] in a network with time-dependent travel times [5]. Many variants of this problem exist. Firstly, paths defined

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

a priori can be distinguished from adaptive decision rules that define a successor based on realized arrival times [5]. We focus on a priori paths, which are more practical than complex decision rules. Secondly, many optimization criteria were proposed, e.g., expected time [5, 6], a combination of expected time and variance [7], worst- [8] and best-case time [9] as well as criteria based on on-time arrival probability [10, 11].

Limiting cost or walking time creates a constrained problem. In its well-studied deterministic version [12] it is NP-complete [13] and can include waiting in some variants [14]. It was recently extended to the stochastic case [15, 16].

Main solution approaches for the unconstrained problem include mixed integer programming [11, 15, 17] and label-correcting algorithms [2, 9, 10, 18]. For example, Chen et al. [2] use an A* search approach to find a shortest path with on-time arrival guarantees and also apply it to a weighted combination of time and cost.

The constrained problem has been addressed with mixed integer programming to minimize expected time [15] and with genetic algorithms to maximize on-time arrival probability [16], both assuming constant weights. Building on Chen et al. [2], our contribution is to develop an exact solution method for the problem with on-time arrival guarantees, applicable to constant as well as time-dependent and stochastic weights.

3 Problem Definition

We use a weighted multi-layer digraph $G = (N, E, L, \Omega, D, W)$ to model intermodal networks. Locations are represented by nodes $i \in N$ and passengers can travel between locations on directed edges $(i, j) \in E \subseteq N \times N$. Each node belongs to a layer $l \in L$ that corresponds to a transport mode, e.g., walking or a train line. Edges within a layer represent travel using that mode and edges between layers model transfers.

Departures and arrivals at nodes are assumed to happen at discrete times t within a finite time horizon $\Omega = \{0, 1, \dots, t_{max}\}$ and travel times are described by discrete random variables. To reflect time-dependence, functions $D_{ij}(t)$ assign such random variables to each edge (i, j) depending on the time $t \in \Omega$ at which it is used. In addition, each edge is labeled with time-dependent discrete random variables $W_{ijk}(t), k \in \{1, \dots, K\}$ representing a diverse set of non-negative weights such as distance or (time-dependent) fees.

Finding robust routes translates into a discrete and stochastic optimization problem:

$$\max_{p, t_{start}} \quad t_{start} \tag{1}$$

$$s.t. \quad P\left(T_d^p(t_{start}) \leq t_{target}\right) \geq \alpha \tag{2}$$

$$P\left(C_{dk}^p(t_{start}) \leq c_k\right) = 1, \forall k \in \{1, \dots, K\} \tag{3}$$

$$t_{start} \in \Omega, \;\; p \in SP_{od} \tag{4}$$

The goal is to find a path p in the set of simple paths SP_{od} from origin $o \in N$ to destination $d \in N$ that leaves the origin at the latest possible time t_{start} and reaches the destination before or at a given time $t_{target} \in \Omega$ with a probability of at least α while satisfying additional constraints on, e.g., the number of transfers, cost or walking time.

The random variable $T_i^p(t)$ in (2) describes the arrival time at node i if starting at time t along path $p = (o, \dots, i, j, \dots d)$. With $T_o^p(t) = t$, it is calculated as follows:

$$T_j^p(t) = T_i^p(t) + D_{ij}(T_i^p(t)), \quad \forall j \in p \backslash o \tag{5}$$

In (3), accumulated weights, described by time-dependent random variables $C_{ik}^p(t)$, $\forall k \in K$ are limited to c_k. With $C_k^p(t) = 0, \forall k \in K$, they are calculated as follows:

$$C_{jk}^p(t) = C_{ik}^p(t) + W_{ijk}(T_i^p(t)), \quad \forall j \in p \backslash o, \forall k \in \{1, \dots, K\} \tag{6}$$

The sums of random variables in (5) and (6) can be calculated using time-dependent convolution. For example, the probability of reaching node j on path p at time w is calculated as follows:

$$P\big(T_j^p(t) = w\big) = \sum_{v \in \Omega, v \leq w} P\big(T_i^p(t) = v\big) P\big(D_{ij}(v) = w - v\big), \forall\, w \in \Omega \tag{7}$$

This requires a sufficiently large time horizon and the assumption of independent edge travel times along paths. However, for scheduled modes, travel times are likely dependent. For example, a train that is delayed on one trip segment might try to get back on schedule on the subsequent segment by driving at top speed. To reflect such dependencies and still maintain the independence assumption, we represent entire journeys on one transportation line with a single edge. The travel time on that edge includes waiting before boarding and total time on board.

Waiting at nodes is often not allowed [2, 18]. However, it can improve the solution in the constrained problem: for example, waiting for congestion charges to drop after rush hour can allow passengers with a limited budget to use a faster mode of transport and thus accelerate their journey. We call this *strategic* waiting, as it is voluntary and thus distinguished from *transfer* waiting, i.e., waiting for a scheduled vehicle. We introduce strategic waiting by replacing $T_i^p(t)$ in (5) and (6) with a random variable $S_i^p(T_i^p(t), s_i)$ representing arrival time after strategic waiting. It depends on the original arrival time $T_i^p(t)$ and waiting parameter $s_i \in \Omega, \forall i \in N$. s_i is an additional decision variable and models that the passenger waits until s_i if arriving before s_i.

4 Solution Approach

We will build on the work of Chen et al. [2], who iteratively solve the closely related forward problem in which t_{start} is fixed and t_{target} is minimized. After several iterations with different starting times, a solution to the original problem is found.

The forward problem will be converted to a multi-objective problem by turning the weight constraints into objective functions. The set of non-dominated solutions to this associated multi-objective problem contains a solution to the constrained problem [19].

To find the non-dominated solutions, we will generalize the A* approach in [2] and apply multi-objective A* search [3]. A* search is defined for additive attributes, but can be extended to convolutions if they are order-preserving [20]. Given two paths p_1, p_2 from o to a node $i \in N \backslash o$ with p_1 dominating p_2 and any path p_3 starting at i, the order-preserving property (OPP) implies that the extended path $p_1 p_3$ dominates the extended path $p_2 p_3$. Given the OPP, dominated paths can be pruned in the algorithm.

Without constraints, the OPP holds if distributions are FIFO[1] [2]. As this is not always true with constraints and general weight functions, we will define conditions for which the OPP holds and structure our proofs along three hypotheses:

- The OPP holds for FIFO travel times and non-decreasing, deterministic weights.
- For deterministic weights that are non-decreasing until time t' and decrease afterwards (e.g., a congestion charge during rush-hour), strategic waiting parameters can be set such that the combined waiting and travel times satisfy the OPP.
- The OPP holds for time-dependent stochastic weights if $t' > t$ implies for all edges that $W_{ijk}(t) \leq W_{ijk}(t'), \forall k \in K$ regarding first-order stochastic dominance (FSD)[2].

5 Next Steps and Expected Contributions

So far, this research in progress has defined the constrained time-dependent reliable shortest path problem with time-dependent and stochastic weights to model robust route planning in urban intermodal traffic. In addition, we outlined an exact solution method.

We will continue to implement this approach and support it with required proofs. Numerical experiments will be conducted to evaluate the algorithm's speed given different network sizes and problem setups. This will include a comparison with the less complex unconstrained problem that was solved within seconds for real-world instances [2]. Furthermore, solutions will be benchmarked against deterministic routing advice. Finally, we plan to set up a showcase for an actual city with travel times sampled from, e.g., real-time public transport data or car trajectories. We expect that this will help to answer important questions for policy-makers such as:

- Which measures are most effective for improving reliability? For example, how does the effect of increasing train reliability compare to that of higher trip frequencies?
- How does reliability affect mode choice? And can mode choice be effectively influenced by improving the reliability of certain modes (e.g., bus or bike sharing)?

By answering these questions, we hope to make a significant contribution to smarter urban mobility in the future.

References

1. Li, Z., Hensher, D.A., Rose, J.M.: Willingness to Pay for Travel Time Reliability in Passenger Transport: A Review and Some New Empirical Evidence. Transp. Res. Part E Logist. Transp. Rev. 46, 384–403 (2010)

---

[1] At any confidence level, travelers arrive at the end of a link in order of departure.

[2] Given discrete random variables X and X' with cumulative distribution functions F_X and $F_{X'}$: X dominates X' regarding FSD $\Leftrightarrow \forall x: F_X(x) \geq F_{X'}(x)$ and $\exists x: F_X(x) > F_{X'}(x)$ [21]

2. Chen, B.Y., Lam, W.H.K., Sumalee, A., Li, Q., Tam, M.L.: Reliable Shortest Path Problems in Stochastic Time-Dependent Networks. J. Intell. Transp. Syst. 18, 177–189 (2014)
3. Mandow, L., Pérez de la Cruz, J.L.: Multiobjective A* Search with Consistent Heuristics. J. ACM. 57, 27:1-27:25 (2010)
4. Frank, H.: Shortest Paths in Probabilistic Graphs. Oper. Res. 17, 583–599 (1969)
5. Hall, R.W.: The Fastest Path through a Network with Random Time-Dependent Travel Times. Transp. Sci. 20, 182–188 (1986)
6. Fu, L., Rilett, L.R.: Expected Shortest Paths in Dynamic and Stochastic Traffic Networks. Transp. Res. Part B Methodol. 32, 499–516 (1998)
7. Sen, S., Pillai, R., Joshi, S., Rathi, A.K.: A Mean-Variance Model for Route Guidance in Advanced Traveler Information Systems. Transp. Sci. 35, 37–49 (2001)
8. Yu, G., Yang, J.: On the Robust Shortest Path Problem. Comput. Oper. Res. 25, 457–468 (1998)
9. Miller-Hooks, E.D., Mahmassani, H.S.: Least Possible Time Paths in Stochastic, Time-Varying Networks. Comput. Oper. Res. 25, 1107–1125 (1998)
10. Nie, Y. (Marco), Wu, X.: Shortest Path Problem Considering On-Time Arrival Probability. Transp. Res. Part B Methodol. 43, 597–613 (2009)
11. Cao, Z., Guo, H., Zhang, J., Niyato, D., Fastenrath, U.: Finding the Shortest Path in Stochastic Vehicle Routing: A Cardinality Minimization Approach. IEEE Trans. Intell. Transp. Syst. 17, 1688–1702 (2016)
12. Di Puglia Pugliese, L., Guerriero, F.: A Survey of Resource Constrained Shortest Path Problems: Exact Solution Approaches. Networks. 62, 183–200 (2013)
13. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York (1979)
14. Desaulniers, G., Villeneuve, D.: The Shortest Path Problem with Time Windows and Linear Waiting Costs. Transp. Sci. 34, 312–319 (2000)
15. Wang, L., Yang, L., Gao, Z.: The Constrained Shortest Path Problem with Stochastic Correlated Link Travel Times. Eur. J. Oper. Res. 255, 43–57 (2016)
16. Li, Y., Guo, L.: Multi-Objective Optimal Path Finding in Stochastic Time-Dependent Transportation Networks Considering Timeliness Reliability and Travel Expense. In: 2016 Prognostics and System Health Management Conference (PHM-Chengdu). pp. 1–6. IEEE (2016)
17. Yang, L., Zhou, X.: Constraint Reformulation and a Lagrangian Relaxation-Based Solution Algorithm for a Least Expected Time Path Problem. Transp. Res. Part B Methodol. 59, 22–44 (2014)
18. Prakash, A.A.: Pruning Algorithm for the Least Expected Travel Time Path on Stochastic and Time-Dependent Networks. Transp. Res. Part B Methodol. 108, 127–147 (2018)
19. Klamroth, K., Tind, J.: Constrained Optimization Using Multiple Objective Programming. J. Glob. Optim. 37, 325–355 (2007)
20. Mandow, L., Pérez de la Cruz, J.L.: Multicriteria Heuristic Search. Eur. J. Oper. Res. 150, 253–280 (2003)
21. Miller-Hooks, E., Mahmassani, H.: Path Comparisons for a Priori and Time-Adaptive Decisions in Stochastic, Time-Varying Networks. Eur. J. Oper. Res. 146, 67–82 (2003)

Track 3:

Unternehmensmodellierung & Informationssystemgestaltung

Track Chairs:

Prof. Dr. Peter Fettke
Universität des Saarlandes / DFKI

Dr. Sebastian Schlauderer
Otto-Friedrich-Universität Bamberg

Work System Modeling Method with Different Levels of Specificity and Rigor for Different Stakeholder Purposes

Steven Alter[1] and Dominik Bork[2]

[1] School of Management, University of San Francisco, San Francisco, USA
alter@usfca.edu
[2] University of Vienna, Faculty of Computer Science, Vienna, Austria
dominik.bork@univie.ac.at

Abstract. This paper proposes a modeling method (the work system modeling method - WSMM) that addresses key issues related to enterprise and process modeling. Those issues lead to modeling method requirements that call for relaxing common assumptions about the nature of modeling methods and related modeling languages and metamodels. A summary of work system theory (WST) and the work system method (WSM) provides background for understanding WSMM. A design space for modeling positions most applications of WSM in relation to seven purposes of modeling that call for successively more formal approaches. WSMM is presented in relation to the seven purposes, thereby extending WSM in new directions. A final section summarizes how WSMM addresses the issues and requirements from the introduction, explains how coherence is maintained within WSMM, and identifies areas for future research.

Keywords: modeling method, modeling language, work system method, work system theory

1 Overcoming Problems in Enterprise and Process Modeling

Prominent researchers argue from various backgrounds that modeling methods for enterprise and process modeling have not achieved their full potential and need to be extended or augmented to make them more usable by broader user groups. This paper's approach to modeling methods addresses important issues that they mention:

- A Dagstuhl seminar [1] emphasized how differing stakeholder needs call for different approaches to enterprise modeling (EM). That seminar led to a *BISE* research note by EM community leaders [2] that encourages moving EM from an expert discipline towards "grass roots modeling" and "modeling for the masses." Their future research agenda includes "softened requirements to completeness, coherence and rigor".
- Six of seven process modeling problems discussed in [3] are relevant here: 1) aiming for one model that suits all purposes, 2) straightjacketing smaller interactive processes into one monolithic model, 3) using static hierarchical decomposition as the only abstraction mechanism, 4) modeling humans as if they are machines doing

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

a single task, 5) being vague about vagueness, 6) abstracting [away] from the things that really matter [to stakeholders].

- [4] calls for "overcoming tendencies to view diagrammatic modeling methods and languages" as "stable, even standardized, artefacts that establish some commonly agreed way of describing a 'system under study,' [which] implies that all stakeholders work on the same level of abstraction and specificity."

Related research on modeling method usage (e.g., [5, 6]) and model comprehension (e.g., [7–9]) illuminates major issues. Many modelers do not apply modeling methods as intended by their designers, frequently using only a subset of the syntactic concepts provided [10]. Modeling methods often do not fit modelers' aptitudes, knowledge, and purposes [11, 12]. [13] notes that the "lack of intuitiveness of diagrammatic representations and the complementary role of text-based representations has been underlined in recent research." Cognitive load [14] for stakeholders becomes increasingly important as unfamiliar symbols and icons proliferate. [13] also mentions lack of flexibility in process models, dilemmas of control, and excessive prescriptiveness. Uncertainty and variability related to accidents, mistakes, and intentional workarounds bring further challenges for modeling methods.

Requirements for a More Flexible Modeling Method. This paper pursues four requirements by presenting a modeling method that relaxes many common assumptions that are obstacles to dealing with those issues about modeling and modeling methods.

1. The modeling method should respect stakeholder diversity related to knowledge, beliefs, and roles, thereby making it usable both by business professionals working individually and in collaboration and by IT professionals pursuing model-driven development or code generation. (cf. [5]).
2. A modeling method can include different modeling techniques for different stakeholder purposes related to the same situation (contrary to the view in [15] that a modeling method can have only one modeling technique that combines a single modeling language and a modeling procedure).
3. With different modeling techniques for different purposes, a modeling method can use different modeling languages based on different metamodels. In relation to domain-specific conceptual modeling (cf., [16]), this approach assumes that intersubjective understanding between stakeholders might not require a single metamodel for processes, services, enterprises, goals, and so on.
4. The representation of a model might or might not use diagrams with rigorously defined notation and syntax (e. g., BPMN, ArchiMate) or might use such diagrams for some purposes but not for others.

Acceptance of multiple techniques, modeling languages, and metamodels within a modeling method leads to challenges about maintaining coherence across different models produced by different stakeholders for different purposes. Our approach to coherence within a modeling method is to require use of a single overarching metaphor that applies to all modeling techniques of the modeling method. According to [17, p. 138], a modeling method metaphor defines a specific perspective taken by the modeler while observing the reality and mapping the relevant aspects to the modeling language at hand, thereby creating a model representation of the reality. Ideally that metaphor should help in bridging gaps between modelers and practitioners who often visualize situations from different viewpoints. The invariance of the single overarching metaphor ensures that all modeling techniques contribute to an overall goal, even if they employ different levels of detail and expressiveness. Differences between the models will be revealed in stakeholders' personal understandings and collaborative discussions.

A Work System Modeling Method. This paper extends several decades of effort related to the work system method (WSM), a flexible systems analysis and design approach based on an informal type of modeling and problem-solving designed to help business professionals visualize work systems and collaborate more effectively with IT professionals [18–21]. Many hundreds of MBA and Executive MBA students have used various versions of WSM templates that guided their production of management briefings about improving real world work systems. Those templates contain many modeling techniques that have never been expressed as a formal modeling language (cf. [22]). For example, none is based on an explicit metamodel or operationalized along a procedure.

Research Goal and Organization. This paper explains how a work system modeling method (WSMM) based on WSM, WST, and a central work system metaphor satisfies the four requirements above. WST and WSM provide a plausible starting point for developing WSMM because their spirit is aligned with the "modeling for the masses" vision in [2] and because an enterprise is a set of interacting work systems.

Building on our previous work [23], this paper provides contributions in several areas by showing that WSMM could help modelers and users apply a range of modeling techniques to situations that seem difficult to address without relaxing assumptions such as use of a single modeling technique, a formal modeling language, and diagrammatic notation. WSMM is a step toward a modeling tool that can be implemented using existing metamodeling platforms such as ADOxx. Reflection on WSMM's form and assumptions might help researchers and practitioners wishing to revise or design modeling methods including those for domain-specific conceptual modeling. It might contribute to reflection on how modeling standards such as BPMN, EPC, and ArchiMate can be adapted to address needs of broader audiences. It also might show a path for formalized extensions of less structured contexts such as those addressed by Checkland's soft systems methodology [24].

The next section summarizes WST and WSM to introduce the central work system metaphor. A two-dimensional design space for modeling methods illustrates the context of WSM. One dimension traverses seven purposes of modeling that require a range of modeling techniques from quite informal to highly formal (cf. [22]). A second dimension represents different degrees of specificity in content, appearance, and usage. The subsequent section illustrates the scope of WSMM using an example that emphasizes modeling techniques and related metamodels. A final section summarizes how WSMM addresses the requirements mentioned above, explains how WSMM maintains coherence across models built for different purposes, and identifies challenges for future research.

2 Thinking of Systems in Organizations as “Work Systems”

This section provides background related to the work system method (WSM), an informal modeling approach that to date has not been guided by a metamodel and that does not produce specifications in the sense of enterprise modeling.

Work system basics. A work system is a system in which human participants and/or machines perform processes and activities using information, technology, and other resources to produce product/services for internal and/or external customers. The and/or in the definition implies that work systems can be sociotechnical (with human participants) or totally automated. A work system operates within an environment that matters (e.g., national and organizational culture, policies, history, competitive situation, demographics, technological change, other stakeholders, and so on). Work systems rely on human, informational, and technical infrastructure that is shared with other work systems. Work systems should support enterprise and departmental strategies. The definition of work system specifies that work system is a very general case that includes many special cases such as information systems, supply chains, service systems, projects, and totally automated work systems. For example, an IS is a work system most of whose activities are devoted to processing information. Supply chains are work systems that extend across multiple organizations to provide resources for other organizations. Projects are work systems that produce specific product/services and then go out of existence. An enterprise or organization is a series of interacting work systems.

WST, the theoretical basis of WSM, consists of three parts: 1) the definition of work system, 2) the work system framework, and 3) the work system life cycle model, which is not discussed here. This paper makes direct use of the definition and of the work system framework (Figure 1), which outlines elements of even a rudimentary understanding of a work system’s form, function, and environment as the work system exists during a time interval when its structure is basically static. Emphasizing business rather than IT concerns, this framework covers situations that might not have a well-defined business process and might not be IT-intensive. Processes and activities, participants, information, and technologies are viewed as completely within the work system. Customers and product/services may be partially inside and partially outside because customers often participate in work systems. A common limit to modeling precision is that human participants in work systems may make errors and may pursue adaptations and workarounds instead of following prescribed procedures. Furthermore, processes fall along a dimension from unstructured to structured [25], starting with largely unstructured creative processes (such as many design and management processes) that have no pre-specified sequence, may involve extensive iteration, and therefore are not amenable to detailed, high precision modeling.

Work system method. WSM is a semi-formal systems analysis and design approach that was developed over several decades to help business professionals visualize work systems in their own organizations and collaborate more effectively with IT professionals. To date, almost all use of WSM has applied work system analysis templates that outline how to proceed from aspects of a work system’s structure and performance toward producing a preliminary recommendation about how to improve the work system. The templates include some questions that require textual answers, others that require filling out formatted tables, and others that invite users to include swimlane diagrams, Pareto charts, or other diagrams if they have appropriate software.

While details differ, every version of WSM is organized around the following: 1) identify the smallest work system that has the problem or opportunity; 2) summarize the “as-is” work system using a work system snapshot (example in Table 2), a stylized one page summary; 3) evaluate work system operation using measures of performance, key incidents, social relations, and other factors; 4) drill down further as necessary; 5) propose changes by producing a work system snapshot of a proposed “to be” work system that will probably perform better; 6) describe likely performance improvements.

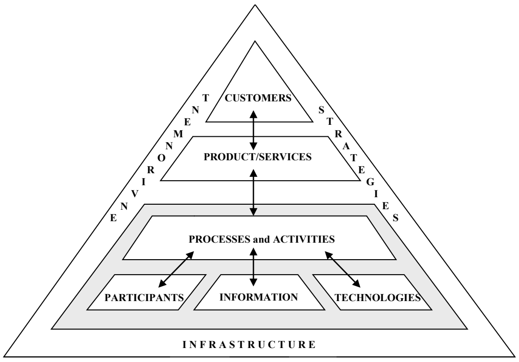


Figure 1. Work system framework [19, 20]

3 Framework for Visualizing Modeling Methods

We agree with the view in [2] that “not all knowledge should be represented as a formal model” and that it is important to find “the right balance of representational forms,” including formal and informal models. A clear discussion of this entire topic requires a foundation such as the framework in [15] by which a modeling method is a composition of a *modeling language*, *modeling procedure*, and *mechanisms & algorithms*. A modeling language, the backbone of a modeling method, is composed of three components: syntax (the concepts provided by a modeling language, including their valid combinations), semantics (the meaning of the concepts), and notation (the graphical representation of the concepts). The combination of a modeling language with a modeling procedure is referred to as a *modeling technique*.

An alternative perspective on modeling methods. Facilitating modeling by diverse stakeholders calls for relaxing the requirement in [15] that all stakeholders and purposes must be accommodated using a single modeling technique, i.e. one modeling language and one modeling procedure. Relaxing that requirement avoids cognitive overload from increasing metamodel complexity due to adding concepts to separate modeling techniques when diverse stakeholder needs and purposes call for them. Thereby, the cognitive load of each single modeling technique is minimized by maintaining the overall expressiveness of the modeling method.

An alternative view of modeling methods starts with a modeling method design space. The ideas explained next are equally applicable to design spaces based on the work system metaphor or other central metaphors such as systems in general, sociotechnical systems, actor networks, activity theory, and viable systems.

3.1 Design Space for Modeling Methods and Modeling Techniques

Figure 2 represents a design space for modeling methods and modeling techniques related to a core metaphor. A key goal of the design space is accommodating a range of stakeholder purposes, shown as P1 through P7. Technique specificity is the extent to which a technique defines exactly what to include, what to ignore, and how to proceed. Techniques with low specificity tend to be flexible but at the cost of providing relatively little conceptual or procedural guidance. The reverse applies as well.

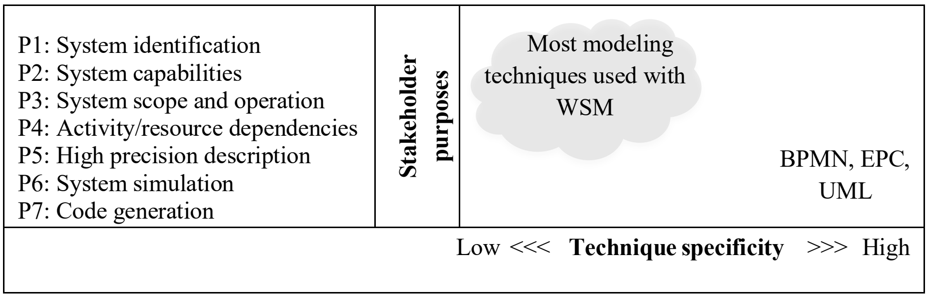


Figure 2. Design Space for Modeling Methods and Modeling Techniques

The shaded area represents the positioning of most of the modeling techniques that WSM users have applied. Most of those techniques focus on topics such as work system *scope and operation*, and *activity/resource dependencies*. Those techniques are relatively low in specificity compared to techniques that might be used for high precision description, system simulation, or code generation. Instead of accepting the assumption that WSMM would include only one technique, this paper assumes that WSMM could include techniques anywhere in the design space provided that those techniques genuinely fit with the overarching metaphor and stakeholder purposes.

4 Work System Modeling Method

WSMM expands WSM greatly by recognizing a wide range of stakeholder purposes and different degrees of specificity for different stakeholders that all use the central work system metaphor. WSMM provides a broader scope of modeling to help various users understand the situation and decide how to improve it. WSMM uses simpler metamodels for informal and intuitive visualization of work systems and more precise and expressive metamodels for helping decision makers identify and select among possible changes and for helping software developers produce or improve software.

WSMM is a highly adaptable method organized around a set of modeling techniques based on a work system metaphor. That range of techniques should be available where required by a range of stakeholders who may encounter situations that resist being modeled within restrictions that are often viewed as essential to conceptual modeling. Our WSMM test case is a hiring system. Similar test cases could be constructed for work systems in production, sales, accounting, software development, and so on.

Managers of an engineering firm are concerned that their current work system for finding, interviewing, and hiring job applicants takes too long, wastes too much effort in interviews, and seems to hire too many engineers whose contributions to engineering projects are disappointing. Some managers believe that a better online human resources (HR) portal would help. Others believe that the problems lie elsewhere.

4.1 Modeling Techniques within WSMM

WSMM should include modeling techniques that support different stakeholder purposes identified in Figure 2. The following discussion uses the hiring example to illustrate the scope of an initial version of WSMM in relation to purposes P1 through P7, but not to show all imaginable modeling techniques that might be included eventually, and certainly not to go into a lengthy explanation of specific techniques.

P1: Identification of the work system. Simply naming the work system of interest with a verb phrase such as *"finding, interviewing, and hiring applicants"* often avoids confusion with people thinking that the technology (e.g., the online HR portal) is the primary object of the improvement effort. Clarity in that regard makes sure that the project is viewed as much more than a software project that produces software. P1 makes no attempt to describe the work system's behavior and structure and does not call for a specific procedure. The metamodel consists of one concept: *Work System.*

Table 1. Two illustrations of lists of capabilities (P2) for the hiring example

Table 1a: Simple list of capabilities	Table 1b: List of capabilities with performance or service level expectations
• defining parameters of the position • publicizing the position • prioritizing applications	• defining parameters for a position, responding within 3 days of request • publicizing the position, advertising in at least five bulletin boards • prioritizing applications, responding within three days of due date

P2: Capabilities of the work system. The hiring work system has capabilities that can be included as a service catalog explanation for the P1 work system description. Those capabilities might be described using a list of verb phrases such as those in Table 1a. A slightly more complete description of capabilities might include performance or service level expectations for each capability in Table 1b. A minimalist metamodel for P2 includes the concepts *Work System* and *Capability* and a slightly more detailed metamodel would add *Performance expectation* or *Service level.* The P2 capabilities lists in Table 1 are simple in form. Again, a specific procedure is not employed.

Researchers focusing on capability driven-development have used a more rigorous notion of capability, e.g., the 15 concepts in the capability-related metamodel in [26].

P3. Scope and general operation of the work system. With P3, the stakeholder wants to clarify the scope of the work system (based on P1) and to attain an overview of its general operation without going into great detail. Table 2 illustrates P3 in the form of a "work system snapshot", a modeling technique for describing a work system's scope and general operation. This is a formatted one-page summary of a work system in terms of the six central elements of the work system framework. Those six elements provide an easily used description that helps in defining the boundaries and contents of the work system that has the problem or opportunity at hand.

Table 2. Illustration related to P3: A work system snapshot of a hiring work system

Customers	Product/Services
• Hiring manager • Larger organization (which will have the applicant as a colleague) • HR manager (who will analyze the nature of applications)	• Applications (which may be used for subsequent analysis) • Job offers • Rejection letters • Hiring of the applicant

Major Processes and Activities	
• **Hiring manager** submits request for new hire within existing budget • **Staffing coordinator** defines the parameters of the new position. • **Staffing coordinator** publicizes position. • **Applicants** submit job applications. • **Staffing coordinator** selects shortlisted applicants. • **Hiring manager** identifies applicants to interview.	• **Staffing coordinator** sets up interviews. • **Hiring manager** and **other interviewers** perform interviews. • **Hiring manager** and **other interviewers** provide feedback from the interviews. • **Hiring manager** makes hiring decisions. • **Staffing assistant** sends offer letters or rejections. • Successful **applicant** accepts or rejects job offer or negotiates further.

Participants	Information		Technologies
• Hiring managers • Staffing coordinator • Applicants • Staffing assistant • Other interviewers	• Job requisition • Job description • Advertisements • Job applications • Cover letters • Resumes	• Short list of applicants • Information and impressions from the interviews • Job offers • Rejection letters	• HR portal for communicating with applicants • Word processor • Telephones • Email

Work system snapshots increase syntactic expressiveness through a metamodel with seven concepts: *Work System, Customer*, *Product/Service*, *Activity*, *Participant*, *Information*, and *Technology*. Cardinalities in the metamodel express internal consistency rules for the snapshot, e.g., each product/service must be received and used by at least one customer group. The type of one-page tabular representation in Table 2 helps in visualizing a work system's scope by focusing on its core components. There is no need for a modeling procedure for producing this type of table. Note that customers may be work system participants, as in custom software development.

P4. Resources used and produced by specific activities. While a work system snapshot such as Table 2 is useful for discussing a work system's purpose and scope, many stakeholders need a deeper understanding of which resources are used and produced by each activity. Tables in the form of Table 3 are more useful for clarifying operational details by listing selected activities of a work system snapshot along with selected types of resources that are used and/or produced by those activities.

Table 3. Illustration of P4: Selected resources used by a subset of Table 2 activities

Activity	Actors	Information used, created, updated, or deleted	Technology	Trigger	Pre-conditions	Post-conditions and product/ services produced
Submit request for new hire.	Hiring manager	Hiring budget Job requisition	HR portal	Need for new employee	Sufficient hiring budget	Job requisition exists
Define parame-ters of the job.	Staffing coordina-tor	Job requisition Job description Hiring policies	Word processor, HR portal	Job requisi-tion	Job requisition	Job description
Publicize the job opening	Staffing coordina-tor	Experience with advertising media, Advertisement	HR portal, Web site for selected media	Job requisite-ion, Job descript-tion	Job requisition, Job description	Advertisement displayed on websites
Submit applica-tion	Applicant	Job description, Cover letter, Job application, Resume	HR portal	Advertise ment displayed on websites	Advertiseme nt displayed on websites	Receipt of cover letter, job application, and resume

Various metamodels can be the basis of Table 3. A minimalist metamodel for P4 would include *Work System, Activity, and Resource*. It would treat all but the first column in Table 3 as resources that are used by the activities. Saying that actors are "used" by activities may sound strange, but it fits with the way some managers use the term resource in planning and management. For modeling this provides a symmetrical way to handle human, informational, and technological resources. A more expressive metamodel might include all of the column headings in Table 3. An even more expressive metamodel presented in [27] contains over 50 concepts that identify other associations between activities and resources, such as which business rules affect customer-facing activities. Those increasingly elaborate metamodels move toward the level of specificity that programmers need, e.g., in identifying resources that are used or produced. The resulting models are still somewhat informal since the concepts are introduced with natural language.

P5. High precision description of the work system. Existing diagrammatic modeling techniques address many typical needs for understanding how the work system components are structured and how the work system operates. For example, it is easy to represent activity sequence and branching logic using BPMN diagrams with activities in swimlanes for different participant roles (see Figure 3, which also applies to P6). While analysts might prefer a full version of BPMN, business stakeholders might prefer a simpler, restricted version of BPMN based on a minimalist metamodel whose concepts are limited to *Event*, *Activity*, *Gateway*, and *Sequence Flow*. That would suffice for diagramming activities in Tables 2 and 3 even though it would require implicit handling of branching logic for situations such as when an applicant is rejected. Similarly, entity-relationship diagrams and ArchiMate's application and technology layers could be used for P5 level descriptions of data structures and interactions between hardware and software. Thus, a high precision description of the work system needs to use modeling techniques that are not associated with WSM or WST but would be important to include in WSMM if it is to address needs of P5, P6, and P7.

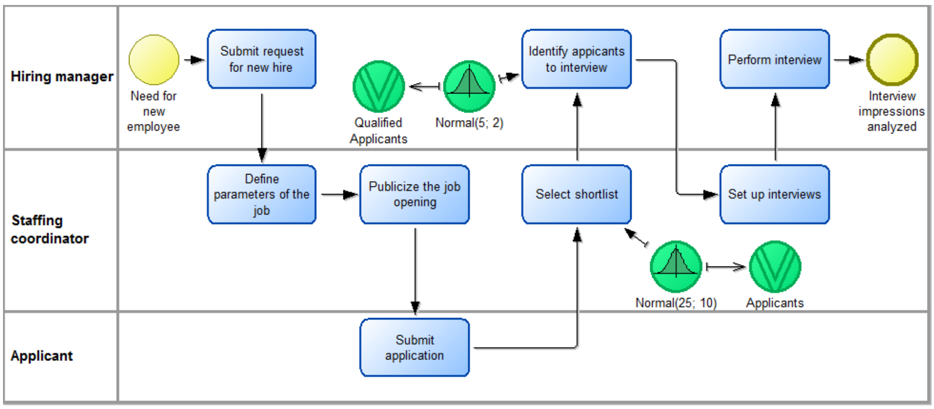


Figure 3. Part of a model supporting a P6 simulation of the hiring process

P6. Simulation of the work system's key processes. Some stakeholders may want to execute simulations to support deeper analysis of the "as-is" work system and deciding on the best of several possible future "to-be" work systems. For example, it might be useful to simulate the workload of different actors depending on the number of applicants, number of interviewers, and other factors. A simulation model could apply an extended BPMN metamodel that adds simulation-specific concepts such as *Statistical Distribution*, and *Random Generators* (the green elements in Figure 3), plus attributes such as transition conditions, probabilities, quantity, cost, and time. Figure 3 illustrates how those concepts might be added to a BPMN model for the hiring example.

P7. Code generation. Figure 2 mentioned code generation because it is a central concern of many researchers in the modeling community. WSMM supports the effort to create understandings and artifacts that are needed in model-driven development but does not attempt to bridge the final gap between understandings and code.

5 Discussion and Conclusions

This paper was inspired by recently published concerns of leading researchers regarding current limitations of enterprise and process modeling. Its goal was to characterize a practical modeling method that applies a central work system metaphor and can be used by stakeholders with different purposes and different levels of technical expertise. WSMM's approach to modeling methods relaxes widely accepted assumptions concerning the nature of modeling methods. That approach positions modeling methods and modeling techniques in a design space that traverses seven types of purposes that had not been articulated in that manner previously and that assumes different levels of technique specificity might be applied for any of those purposes.

We conclude with three topics: whether the four requirements for WSMM were met, how coherence is maintained, and areas for future research.

5.1 Satisfying Four Requirements for WSMM

WSMM satisfies the four requirements identified as the outset. 1) In contrast with most modeling methods, it respects stakeholder diversity by recognizing that different stakeholder purposes may generate different needs for expressiveness, rigor, and completeness. 2) It includes multiple modeling techniques, as shown in the previous section. 3) It can use multiple modeling languages based on different metamodels, but all related to the same work system metaphor. 4) It includes diagrams where needed by stakeholders, e.g., for P5 and P6, but does not require diagrams for other purposes.

Similarly, WSMM addresses other issues mentioned at the outset. WSMM conforms with the willingness in [2] to soften requirements for completeness, coherence, and rigor to achieve broader and more effective usage of modeling. In relation to process modeling problems discussed in [3]: 1) WSMM assumes that different users with different purposes will prefer models with different characteristics; 2) WSMM does not straitjacket processes into one monolithic model; 3) WSMM does not use static decomposition; 4) WSMM does not treat people as machines doing a single task; 5) WSMM is clear about vagueness in its recognition of unstructured and semi-structured processes, adaptations, and workarounds; 6) WSMM is designed to help people improve work systems and does not abstract away from things that matter, such as performance. In relation to issues from [4], WSMM assumes that stakeholders will work at different levels of abstraction and specificity and that their needs may not be satisfied by a specific modeling language.

5.2 Coherence within WSMM

The challenge of coherence in WSMM can be viewed as making sure that WSMM is more than just an assemblage of techniques. At minimum, the use of metamodels that all build on the same core metaphor (the work system) should facilitate production of consistent models for different purposes. For example, a P2 model of capabilities should create a logical constraint on a P3 model that summarizes the work system scope and operation that enacts those capabilities. P3 processes and activities become the basis for the activities in P4 models, and so on.

A larger question is whether WSMM would help business professionals understand work systems for their own purposes while also helping them collaborate with IT professionals with different world views. The use of different metamodels developed for different purposes reduces the likelihood of an automatic way to zoom between the various modeling techniques. However, the progression of the different stakeholder purposes (P1 through P7) outlines a path for communicating between stakeholders who have different purposes. P1 names the work system in the form of the verb phrase. All stakeholders should be able to rally around creating or improving that work system. P2 and P3 support relatively informal understandings of the work system's content, capabilities, and scope. A software engineer in a work system improvement project needs to understand levels P1, P2, or P3 to communicate effectively with business professionals about topics other than isolated details, and also to make it less likely that software will miss basic issues. P4 covers details that business professionals need to verify and that IT professionals need to understand to produce software. Both business and IT professionals might have their own needs for delving into detailed process models, data models, or technical interface models created for P5. In all cases, knowledge of models for P1 through P4 will help in understanding both details and significance of models for P5. Only specialists will pursue P6 and P7, and it would be difficult for them to do that well without understanding models for P3 and P4.

Much of coherence across the various modeling techniques in WSMM needs to be achieved through looking at models and communicating with other stakeholders. From a general standpoint, the different techniques are completely independent and their purposes might not follow the linear and successive nature as proposed for WSMM. Therefore, a technical realization is not always necessary or feasible and needs to be decided individually for each modeling method and its modeling techniques.

5.3 Areas for Future Research

WSMM could help modelers and users apply a range of valuable modeling techniques to situations that seem difficult to address convincingly without relaxing widely held assumptions about the singular nature of modeling methods and modeling languages. Here are some of the research opportunities that represent next steps:

Exploring general implications of the WSMM vision. This paper showed how relaxing common assumptions about modeling methods enables comprehensive visualization of complex systems such as work systems. We do not know where synergies and conflicts between this view and more established views might take us.

Challenging, extending, and/or validating these ideas requires iterations of discussion, feedback, and possibly revision.

Implementation of a widely accessible WSMM toolkit. This paper focused on explaining ideas rather than illustrating technical implementations. In a parallel effort we demonstrated technical feasibility of implementing WSMM by using ADOxx, a widely used metamodeling development and configuration platform for implementing modeling methods that is available through OMiLAB (the Open Models Laboratory), an open community for the conceptualization of modeling methods [28, 29]. ADOxx provides capabilities that can be used to implement all of the metamodels mentioned in this paper plus all of the modeling techniques found in existing Microsoft Word analysis templates used by MBA and EMBA students in recent years. For example, we created a blank work system snapshot in ADOxx and filled in the details from Table 2. ADOxx also can support other work system modeling techniques such as those related to conformance to sociotechnical principles, anticipation of workarounds, customer responsibilities for specific work system activities, and the value of product/services to specific customer groups. Further work and experience with ADOxx is required to identify the most convenient ways to use its modeling capabilities across the many topics in existing WSM templates and other topics based on new metamodels.

Replication for other test cases. Many aspects of this paper, such as the dimensions of the design space in Figure 2, superimposed a rigorous modeling viewpoint on top of useful but less rigorous ideas associated with WSM as it has existed to date. The same type of exercise should be attempted for other sets of ideas, such as general systems theory, sociotechnical theory, actor network theory, soft system methodology, and so on. That would require experts in those areas of theory and practice to identify modeling techniques that are used or could be used, to specify underlying metamodels, and to explore the possibility of producing modeling methods that researchers and practitioners in those areas would find useful.

Further development of the research stream related to WSM, WST, and extensions. The development of WSM started several decades ago with a focus on issues related to P3 and with little or no attempt at rigor other than trying to define terms and encourage organized thinking about work systems that involved IT. The ideas that became WSM and WST evolved gradually. A first book on WSM appeared in 2006; a first version of a work system metamodel in 2010 as an extension of the core ideas; a first article articulating WST appeared in 2013.

The current research extends that stream of research in many directions. It overcomes the limiting assumption that the research was mostly centered around what Figure 2 would call a P3 analysis by business professionals. It eliminates an outdated assumption of a single work system metamodel that needs to be highly detailed (and that until recently was updated incrementally). The new approach is potentially much more valuable because it calls for many alternative metamodels based on the same central metaphor but designed for different purposes.

This paper described metamodels for six of the seven purposes in Figure 2 and noted that different metamodels with different degrees of specificity could be applied for most of those purposes. Length limitations prevented including diagrams of the metamodels. Follow-on research should look at work system-related metamodels in substantial detail, showing them as diagrams, and most importantly, explaining the stakeholder issues that would call for greater syntactic and semantic expressiveness related to specific issues. That research would contrast incremental extensions at each level based on well-known stakeholder concerns (e.g., related to different types of information, product/service offerings, value for customers, application of encapsulated services, and various types of interactions with other work systems) with the combination of more unstructured relationships among modeling techniques.

References

1. Clark, T., Frank, U., Kulkarni, V.: Supporting Organizational Efficiency and Agility: Models, Languages and Software Systems, https://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=16192.
2. Sandkuhl, K., Fill, H.-G., Hoppenbrouwers, S., Krogstie, J., Matthes, F., Opdahl, A., Schwabe, G., Uludag, Ö., Winter, R.: From Expert Discipline to Common Practice: A Vision and Research Agenda for Extending the Reach of Enterprise Modeling. Bus. Inf. Syst. Eng. 60, 69–80 (2018).
3. van der Aalst, W.M.P.: What makes a good process model? Softw. Syst. Model. 11, 557–569 (2012).
4. Karagiannis, D.: Agile modeling method engineering. In: Karanikolas, N.N., Demosthenes, A., Mara, N., Vergados, D., and Michalis, X. (eds.) Proceedings of the 19th Panhellenic Conference on Informatics - PCI '15. pp. 5–10. ACM Press, Athens, Greece (2015).
5. Fettke, P.: How Conceptual Modeling Is Used. Commun. Assoc. Inf. Syst. 25, 571–592 (2009).
6. Mendling, J., Reijers, H.A., van der Aalst, W.M.P.: Seven process modeling guidelines (7PMG). Inf. Softw. Technol. 52, 127–136 (2010).
7. Haisjackl, C., Soffer, P., Lim, S.Y., Weber, B.: How do humans inspect BPMN models: an exploratory study. Softw. Syst. Model. 17, 655–673 (2018).
8. Johannsen, F., Leist, S., Braunnagel, D.: Testing the Impact of Wand and Weber's Decomposition Model on Process Model Understandability. ICIS 2014 Proc. (2014).
9. Mendling, J., Recker, J., Reijers, H.A., Leopold, H.: An Empirical Review of the Connection Between Model Viewer Characteristics and the Comprehension of Conceptual Process Models. Inf. Syst. Front. (2018).
10. Langer, P., Mayerhofer, T., Wimmer, M., Kappel, G.: On the Usage of UML: Initial Results of Analyzing Open UML Models. In: Fill, H.-G., Karagiannis, D., and Reimer, U. (eds.) Modellierung. p. 21 (2014).
11. Muehlen, M. zur, Recker, J.: How Much Language Is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In: Seminal Contributions to Information Systems Engineering. pp. 429–443. Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
12. Hinkel, G., Kramer, M., Burger, E., Strittmatter, M., Happe, L.: An empirical study on the perception of metamodel quality. In: 2016 4th International Conference on Model-Driven Engineering and Software Development (MODELSWARD). pp. 145–152 (2016).
13. Simões, D., Antunes, P., Carriço, L.: Eliciting and Modeling Business Process Stories. Bus. Inf. Syst. Eng. 60, 115–132 (2018).
14. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. Learn. Instr. 4, 295–312 (1994).
15. Karagiannis, D., Kühn, H.: Metamodelling Platforms. In: Third International Conference EC-Web 2002. p. 182 (2002).

16. Karagiannis, D., Mayr, H.C., Mylopoulos, J.: Domain- Specific Conceptual Modeling. Springer Berlin Heidelberg (2016).
17. Ferstl, O.K., Sinz, E.J.: Grundlagen der Wirtschaftsinformatik. Oldenbourg, München (2013).
18. Alter, S.L.: How should business professionals analyze information systems for themselves? Presented at the (1995).
19. Alter, S.: The work system method: connecting people, processes, and IT for business results. Work System press, Lankspur, CA, CA (2006).
20. Alter, S.: Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future. J. Assoc. Inf. Syst. 14, 72–121 (2013).
21. Truex, D., Alter, S., Long, C.: Systems analysis for everyone else: Empowering business professionals through a systems analysis method that fits their needs. In: ECIS 2010 Proceedings. 4 (2010).
22. Bork, D., Fill, H.-G.: Formal Aspects of Enterprise Modeling Methods: A Comparison Framework. In: Proceedings of the Annual Hawaii International Conference on System Sciences. pp. 3400–3409. IEEE (2014).
23. Bork, D., Alter, S.: Relaxing Modeling Criteria to Produce Genuinely Flexible, Controllable, and Usable Enterprise Modeling Methods. In: Fellmann, M. and Sandkuhl, K. (eds.) 9th International Workshop on Enterprise Modeling and Information Systems Architectures (EMISA) (2018).
24. Checkland, P.B.: Soft systems methodology. Hum. Syst. Manag. 8, 273–289 (1989).
25. Alter, S., Recker, J.: Using a work system perspective to expand BPM research use cases for research. J. Inf. Technol. Theory Appl. 18, 47–71 (2017).
26. Zdravkovic, J., Stirna, J., Kuhr, J.-C., Koç, H.: Requirements Engineering for Capability Driven Development. In: Frank, U., Loucopoulos, P., Pastor, Ó., and Petrounias, I. (eds.) IFIP Working Conference on The Practice of Enterprise Modeling. pp. 193–207. Springer, Berlin, Heidelberg (2014).
27. Alter, S., Bolloju, N.: A Work System Front End for Object-Oriented Analysis and Design. Int. J. Inf. Technol. Syst. Approach. 9, 1–18 (2016).
28. Bork, D., Miron, E.-T.: OMiLAB - An Open Innovation Community for Modeling Method Engineering. In: International Conference of Management and Industrial Engineering (ICMIE) - Management in the Innovation Society. pp. 64–77. Niculescu Publishing, Bucharest, Romania (2017).
29. Bork, D., Buchman, R.A., Karagiannis, D., Lee, M., Miron, E.-T.: An Open Platform for Modeling Method Conceptualization: The OMiLAB Digital Ecosystem. Commun. Assoc. Inf. Syst. (2019).

Resolving Inconsistencies in Declarative Process Models based on Culpability Measurement

Carl Corea, Matthias Deisen, and Patrick Delfmann

Institute for Information System Research, University of Koblenz-Landau, Germany
{ccorea,mdeisen,delfmann}@uni-koblenz.de

Abstract. Contrary to traditional process models, declarative process models define a set of declarative constraints to specify the behavior which a process should adhere to. In the scope of process mining, declarative process discovery aims to derive such constraint sets from event logs. Here, a problem for current discovery techniques is that of inconsistency. That is, dependent of certain event log characteristics, the derived constraint set may contain contradictory constraints. This in turn however makes the discovered model unusable, as contradictory constraints make it impossible to execute declarative process models, thus hampering previous process discovery efforts. In this work, we present an approach for resolving inconsistencies in declarative process models, based on methods from the scientific field of inconsistency measurement. We introduce our approach algorithm and evaluate its feasibility with data sets of the BPI Challenge 2017.

Keywords: Declarative Process Models, Inconsistency Resolution, Declare

1 Introduction

Process Discovery is a key part of Process Mining and comprises techniques for the automated derivation of process models [18, 20]. The creation of processes models is performed by the means of discovery algorithms and techniques, which are applied to entail process models based on observed behaviour, for example that in event logs [6, 17]. Recent efforts have been directed towards so-called declarative process models [17]. Where procedural process models provide an imperative description of how exactly company processes should be performed, declarative process models consist of a set of constraints, that must not be violated. Hence, contrary to the explicitly confined process execution in procedural process models, process execution in declarative process models is all allowed behaviour within the set of constraints.

As the declarative constraint set defines how processes can be executed, the quality of this set is of central importance to companies, in order to ensure a correct and compliant process execution. A potential problem here is that of inconsistency [6]. That is, the set of constraints must not contain any contradictory constraints, as this can make it impossible to perform the process such that it satisfies the set of declarative constraints, i.e. the declarative process model cannot be executed. Ensuring the

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

consistency of declarative process models in the scope of process discovery is widely recognized as a challenging task [6, 8, 15]. In cases of inconsistency, methods are needed to provide an analysis and (semi-) automated resolution capabilities, as a manual analysis and resolution is not feasible in practice due to the potential size of declarative process models.

In this paper, we describe an approach that can automatically resolve inconsistencies in declarative process models. The approach applies quantitative measures from the scientific field of inconsistency measurement [10]. Such measures allow to assign a numerical value to individual constraints, with the informal meaning that a higher value reflects a higher degree of inconsistency. In result, this quantitative assessment is used to pin-point and delete causes of inconsistency in the declarative process model while ensuring a low amount of information loss. We implemented our approach and conducted an evaluation of data sets of the Business Process Intelligence Challenge 2017.

The remainder of this work is structured as follows. In Section 2, we provide a motivational example and introduce preliminaries. Next, Section 3 introduces our approach algorithm to resolve inconsistencies in declarative process models. The evaluation of our implemented algorithm is presented in Section 4. Last, we discuss our approach in the context of related work in Section 5 and conclude in Section 6. As a design choice, we base our approach on the Declare formalisms, which is a widely acknowledged modelling language for declarative process models [6, 12, 17]. Our approach can however be extended to arbitrary declarative process modelling languages as defined subsequently.

2 Background

This section provides an example for the necessity of resolving inconsistencies in declarative process models. Also, prerequisites for the Declare modelling language and culpability measurement are introduced.

2.1 Declare Modelling Language and Motivational Example

Declarative process models can be defined as a set of constraints, modelling the allowed behaviour that processes should adhere to.

Definition 1 (Declarative Process Model). A declarative process model is a tuple $\boldsymbol{M} = (\boldsymbol{A}, \boldsymbol{T}, \boldsymbol{C})$, where $\boldsymbol{A}$ is a set of *tasks*, $\boldsymbol{T}$ is a set of pre-defined constraint *templates*, and $\boldsymbol{C}$ is the set of actual *constraints*, which instantiate the template elements in $\boldsymbol{T}$ with activities in $\boldsymbol{A}$.

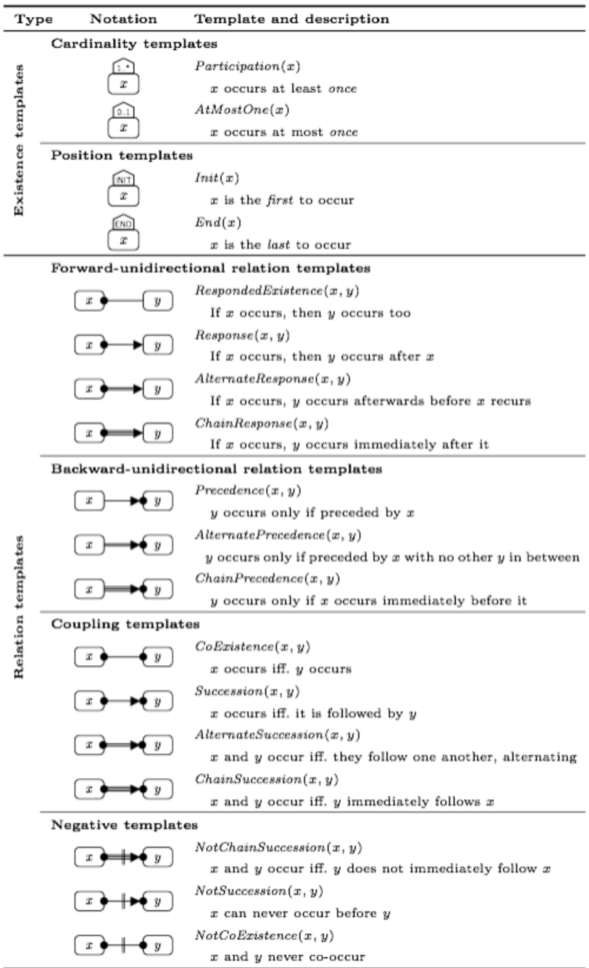
In this paper, we consider Declare, which is a popular declarative modelling language based on a set of constraint templates. Such templates can be used to define declarative constraints by passing tasks as parameters. For instance, the template *init(x)* indicates that every process instance must start with a task x. Declare can be used to define the *relationships* between two individual tasks, by the means of so-called relationship templates, each taking two input parameters. For example, *Response(x,y)* indicates that for any process instance, if *x* occurs, then later *y* must occur. The template *ChainResponse(x,y)* denotes that for any process instance, if *x* occurs, then *y* must occur *immediately* after *x*, i.e., a directly follows relation. Due to space limitations, we ommit a further detailed discussion of all relationship templates. An overview and a short description is provided in Figure 1. Please see [6, 17] for a further discussion.

The template relationships can be hierarchically ordered. Figure 2 shows the subsumption hierarchy of relational Declare templates considered in this work. For example, a *ChainResponse* between tasks *a* and *b* is also a *Response* between *a* and *b*.

Three important templates for the remainder of our discussion are the so-called *negative* relationship templates *NotCoExistence(x,y)*, *NotSuccession(x,y)* and *NotChainSuccession(x,y)*. *NotCoExistence(x,y)* states that tasks *x* and *y* must never appear in the same process instance. *NotSuccession(x,y)* states that *y* must never occur after *x* in a process instance (regardless of how many tasks are in-between *x* and *y*). *NotChainSuccession(x,y)* states that *y* must not *directly* follow *x* for any process instance. These three negative templates can cause inconsistency of the overall declarative process model in combination with other templates. For example, *ChainSuccession(a,b)* and *NotChainSuccession(a,b)* must not appear simultaneously in the same constraint set, as this is a logical contradiction. As the set of Declare templates is pre-defined, it is possible investigate which combinations of templates can cause inconsistencies. Here we distinguish between three types of Declare inconsistencies, namely *trivial inconsistencies*, *generalization-based inconsistencies* and *path-based inconsistencies*.

- **Trivial Inconsistencies.** We define trivial inconsistencies as the co-existence of any negative constraint and its direct complement, i.e. with the same parameters, in the same constraint set. That is, any constraint-set with the templates *CoExistence(x,y)* and *NotCoExistence(x,y)*, OR, *Succession(x,y)* and *NotSuccession(x,y)*, OR, *ChainSuccession(x,y)* and *NotChainSuccession(x,y)* is trivially inconsistent.

- **Generalization-based Inconsistencies.** The subsumption hierarchy entails that generalization relations can also impose inconsistencies in combination with negative templates. For instance, every *ChainSuccession(x,y)* is also a *Succession* between *x* and *y*. Therefore, the two constraints *ChainSuccession(x,y)* and *NotSuccession(x,y)* are contradictory to each other. Based on the subsumption hierarchy shown in Figure 2, all possible combinations of generalization-based inconsistencies can be defined. Intuitively, *NotCoExistence(x,y)* contradicts all other (non-negative) templates with the parameters *x* and *y*, as any possible occurrence of both *x,y* contradicts the *NotCoExistence* of *x* and *y*. *NotSuccession(x,y)* contradicts *Precedence(x,y), Succession(x,y), Response(x,y)* and all inheriting template types. Last, *NotChainSuccession(x,y)* contradicts *ChainPrecedence(x,y), ChainSuccession(x,y)* and *ChainResponse(x,y)*.



Type	Notation	Template and description
Existence templates		**Cardinality templates**
		Participation(x) — x occurs at least *once*
		AtMostOne(x) — x occurs at most *once*
		Position templates
		Init(x) — x is the *first* to occur
		End(x) — x is the *last* to occur
Relation templates		**Forward-unidirectional relation templates**
		RespondedExistence(x, y) — If x occurs, then y occurs too
		Response(x, y) — If x occurs, then y occurs after x
		AlternateResponse(x, y) — If x occurs, y occurs afterwards before x recurs
		ChainResponse(x, y) — If x occurs, y occurs immediately after it
		Backward-unidirectional relation templates
		Precedence(x, y) — y occurs only if preceded by x
		AlternatePrecedence(x, y) — y occurs only if preceded by x with no other y in between
		ChainPrecedence(x, y) — y occurs only if x occurs immediately before it
		Coupling templates
		CoExistence(x, y) — x occurs iff. y occurs
		Succession(x, y) — x occurs iff. it is followed by y
		AlternateSuccession(x, y) — x and y occur iff. they follow one another, alternating
		ChainSuccession(x, y) — x and y occur iff. y immediately follows x
		Negative templates
		NotChainSuccession(x, y) — x and y occur iff. y does not immediately follow x
		NotSuccession(x, y) — x can never occur before y
		NotCoExistence(x, y) — x and y never co-occur

Figure 1. Overview of Declare templates (Taken from [6])

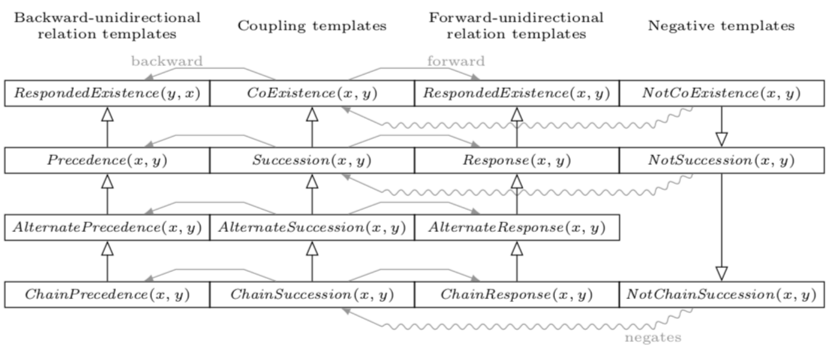


Figure 2. Declare Subsumption Hierarchy (Taken from [6])

- **Path-based Inconsistencies.** So far, we have only discussed the relation of templates with identical parameter signature, e.g. *(x,y)* for both templates. However, *transitive* relations between individual constraints must also be considered. Consider a declarative constraint set consisting of *ChainSuccession(a,b), ChainSuccession(b,c)* and *NotSuccession(a,c)*. The first two constraints state that for every process instance, *a* must directly be followed by *b*. Accordingly, *b* must be directly followed by *c*. Yet, the constraint *NotSuccession(a,c)* demands that *a* must never occur before *c* in the same process instance. Thus, the demands of the first two constraints are contradictory with *NotSuccession(a,c)*. Therefore, inconsistency can occur between groups of templates even if they don't have the same parameters. The definition of path-based inconsistencies will be further discussed in Section 3 based on the definition of so-called *task entailment graphs*.

To clarify, consider the following declarative process model D_1 based on Declare.

ChainSuccession(a,b)	*(i)*	*Response(b,d)*	*(vi)*
NotChainSuccession(a,b)	*(ii)*	*ChainResponse(d,e)*	*(vii)*
NotSuccession(a,b)	*(iii)*	*ChainResponse(e,c)*	*(viii)*
ChainSuccession(b,c)	*(iv)*	*ChainResponse(a,b)*	*(ix)*
NotSuccession(a,c)	*(v)*		

Figure 3. Exemplary Declare model D_1

First, the constraint set in Figure 3 has a trivial inconsistency, as *(i) ChainSuccession(a,b)* and *(ii) NotChainSuccession(a,b)* is contradictory. Also, there are generalization-based inconsistencies. *NotSuccession(a,b)* also entails *NotChainSuccession(a,b)*, therefore, line *(iii)* contradicts line *(i)*. This also yields a conflict between line (iii) and line (ix). Next, the constraints in lines *(i)* and *(iv)* constrain the tasks *abc* as a valid process instance. This path is in contradiction to *(v) NotSuccession(a,c)*, thus a path-based inconsistency. Also, the constraints in line *(i), (vi), (vii)* and *(viii)* define the sequence *abdec* as a valid process instance, which constitutes a path-based inconsistency to *(v) NotSuccession(a,c)*. Path-based inconsistencies also arise for *(ix)*, *(vi), (vii), (viii)* with *(v)*, and (ix), (iv) with (v).

The question may arise how a constraint set such as in Figure 3 can even exist in the first place. Following Di Ciccio et al. (2017), Declare models are mostly automatedly generated in the scope of declarative process discovery. A central problem here are certain completeness characterictics in the underlying event logs. Recording errors or irregularities in process execution can lead to incomplete or distorted logs. To cope with such problems, most declarative process discovery algorithms introduce the notion of a *support factor*, that is, a parameter defining the fraction of traces in which a discovered template must occur in, in order to accept it in the resulting declarative constraint set [6]. Due to the mentioned problems of log completeness, it can make sense to lower this support factor. However, while this increases the amount of constraints, this does not ensure that these constraints are consistent in relation to each other. Thus, methods for post-evaluation in process discovery are needed to ensure that the returned constraint set is consistent.

2.2 Culpability Measurement

A scientific field concerned with the analysis of inconsistent information is the field of Inconsistency Measurement, cf. Grant and Martinez (2018). Here, a central object of study are quantitative measures, which allow to assign a numerical value to (elements of) a constraint set, with the informal meaning that a higher value reflects a higher degree of inconsistency, cf. Thimm (2016) for a survey.

A core notion of our approach are *culpability measures*, which are a type of mentioned quantitative measures [5]. Informally, culpability assesses the degree of blame that an individual constraint carries in the context of the overall inconsistency of the constraint set [11]. Let $\mathfrak{D}$ be the set of all declare constraint sets, and $\mathfrak{C}$ the set of individual constraints in a set $\in \mathfrak{D}$. Then, a culpability measure $\mathcal{C}$ is a function

$$\mathcal{C}: \mathfrak{D} \times \mathfrak{C} \rightarrow [0, \infty) \tag{1}$$

which assigns a non-negative real value to a mapping of a declare constraint set and an individual constraint. An example is the so-called $C_{\#}$ measure [11] which assesses the culpability of a constraint γ for a constraint set $\boldsymbol{D}$ via

$$C_{\#}(\boldsymbol{D}, \gamma) = | \, m \in \boldsymbol{MIS}(\boldsymbol{D}) \mid \gamma \in m \, | \tag{2}$$

This measure is based on so-called minimal inconsistent subsets of the constraint set $\boldsymbol{D}$. Given a constraint set $\boldsymbol{D}$, the minimal inconsistent subsets (***MIS***) of $\boldsymbol{D}$ are defined via

$$\boldsymbol{MIS}(\boldsymbol{D}) = \{\boldsymbol{D'} \subseteq \boldsymbol{D} \mid \boldsymbol{D'} \textit{ is inconsistent and minimal in terms of set inclusion}\} \tag{3}$$

This definition of minimal inconsistent subsets can be used to find inconsistencies in declarative constraint sets. We revisit the exemplary constraint set D_1 from Figure 3. Figure 4 shows the minimal inconsistent subsets of D_1.

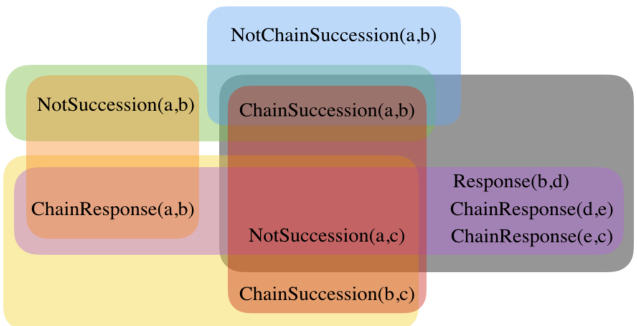


Figure 4. Minimal Inconsistent Subsets for Example D_1 from Figure 3

Based on these ***MIS***, the $C_{\#}$ measure counts the number of minimal inconsistent subsets that a constraint γ belongs to. Applying the $C_{\#}$ measure for the ***MIS*** shown in Figure 4 results in the quantification shown in Figure 5.

C#(D1, ChainSuccesion(a,b))	*= 4*	*C#(D1, Response(b,d))*	*= 2*
C#(D1, NotChainSuccession(a,b))	*= 1*	*C#(D1, ChainResponse(d,e))*	*= 2*
C#(D1, NotSuccession(a,b))	*= 2*	*C#(D1, ChainResponse(e,c))*	*= 2*
C#(D1, ChainSuccession(b,c))	*= 2*	*C#(D1, ChainResponse(a,b))*	*= 3*
C#(D1, NotSuccession(a,c))	*= 4*		

Figure 5. Constraint Culpability for Example D_1 from Figure 3

As a driver for inconsistency resolution, the authors in [5] introduced the notion of a *culpability ranking*, which orders the constraints by their degree of culpability. For the quantification shown in Figure 5, this leads to the following ranking

$$\langle i, v, ix, iii, iv, vi, vii, viii, ii\rangle \tag{4}$$

indicating a prioritization of which constraints should be attended to.

In the following Section, we discuss how this culpability ranking can be exploited for inconsistency resolution and show our algorithm, which includes the computation of minimal inconsistent subsets, culpability assessment and ranking.

3 Inconsistency Resolution Algorithm

3.1 Approach

To resolve inconsistency in a declarative constraint set, there are generally two possibilities [6]: One, a *new* constraint set is constructed, by iteratively moving elements of the old set to the new set, if their introduction does not cause an inconsistency. This technique can be compared to an optimisation problem. A potential disadvantage here is an information loss due to local optima [6]. A different possibility to resolve inconsistency is to use the original constraint set, and iteratively *delete* elements until the constraint set is consistent again. In this context, Grant and Hunter (2011) denote this approach as stepwise inconsistency resolution, and discuss its advantages related to less information loss. In order to resolve inconsistencies with the goal of mitigating information loss, we therefore base our approach on the latter approach of stepwise resolution. Let a constraint set $\boldsymbol{D}$, the deletion of a constraint γ from $\boldsymbol{D}$ is defined via $deletion(\gamma) = \boldsymbol{D} \setminus \{\gamma\}$. Stepwise inconsistency resolution by *deletion* is thus a sequence of deletions $S = <d_1,\ldots,d_n>$. We denote $d_\gamma(\mathbf{D})$ as the constraint set obtained from deleting the element γ from $\boldsymbol{D}$. We denote $\mathbf{\Gamma}_S = < \gamma_1,\ldots, \gamma_n>$ as the individual constraints deleted in the Sequence S via the deletions $d_1,\ldots,d_n$.

The challenge is to find suitable elements which to delete, such that the sequence S results in a low amount of information loss. The authors in [11] do not provide means to determine which elements should be deleted. Here, our algorithm extends the approach by [11] and utilizes the introduced culpability measure $C_{\#}$ to automatedly determine which elements to delete. Let a constraint γ with a $C_{\#}$ value of n_{γ}, then, deleting γ results in the resolution of n_{γ} minimal inconsistent subsets [9]. Thus, we exploit the culpability ranking to find the constraint with the highest culpability, which maximises the number of resolved minimal inconsistent subsets with loosing only one constraint. For example, as can be seen from Figure 4, deleting the constraint *ChainSuccession(a,b)* with the highest culpability value of 4 would result in the deletion of 4 minimal inconsistent subsets.

We iteratively perform the steps of *analysis*, *ranking* and *deletion* until the constraint set is consistent. In the following, we discuss the these steps of our algorithm in detail.

3.2 Algorithm

Our algorithm is shown in Figure 6.

Algorithm 1 Inconsistency Resolution

```
1: procedure RESOLVEINCONSISTENCIES(constraintSet)
2:     mis ← computeMIS(constraintSet)
3:     tempMaxConstraint ← ∅
4:     tempValue ← 0
5:     while mis.size > 0 do
6:         for each c : constraintSet do
7:             culpability ← | m ∈ mis | c ∈ m |
8:             if culpability > tempValue then
9:                 tempMaxConstraint ← c
10:                tempValue ← culpability
11:        constraintSet.delete(tempMaxConstraint)
12:        mis ← computeMIS(ConstraintSet)
```

Figure 6. Approach Algorithm

The algorithm takes as input parameter a set of constraints. A first step is the computation of minimal inconsistent subsets in line 2. (cf. the below discussion). The algorithm then performs the steps of a) analyzing the constraint with the highest culpability (lines 6-10), and b) deleting the constraint with the highest culpability (line 11), in an iterative

```
Algorithm 2 Computation of Minimal Inconsistent Subsets
1: function COMPUTEMIS(constraintSet)
2:     mis ← {∅}
3:     negativeTemplates ← {NotCoExistence, NotSuccession, NotChainSuccession}
4:     negativeConstraints ← getConstraintsInType(negativeTemplates)
5:
6:     for each n : negativeConstraints do
7:         contradictions ← getRulesByTypeAndParams(n.type.complement, n.params)
8:         for each c : contradictions do
9:             mis ← mis ∪ {n, c}
10:
11:    for each n : negativeConstraints do
12:        for each compType : n.type.complementSet do
13:            contradictions ← getRulesByTypeAndParams(compType, n.params)
14:            for each c : contradictions do
15:                mis ← mis ∪ {n, c}
16:
17:    graph = new ConstraintGraph(constraintSet)
18:    for each ns : getConstraintsByType(NotSuccession) do
19:        pathSet ← computeAllPaths(graph, ns.params)
20:        for each p : pathSet do
21:            mis ← mis ∪ {ns, p.getConstraints()}
       return mis
```

Figure 7. Computation of Minimal Inconsistent Subsets in Declare

manner, while there are still inconsistencies (line 5). To clarify, the deletion of constraints intuitively bears the danger of information loss. Yet, deleting constraints might be necessary to restore consistency, i.e. if there exist inherently contradictory constraints. In this context, our approach offers a recommendation as to which element should be deleted, with the goal of maximizing consistency while deleting the lowest amount of constraints possible. We also store the deleted elements and present them to the user for further inspection after the automated resolution process.

A substantial part of our algorithm is the computation of minimal inconsistent subsets, shown in Figure 7. The function depicted in Figure 7 takes as input parameter the declarative constraint set, and returns the set of minimal inconsistent subsets. We initialize an empty set *mis* for storing minimal inconsistent subsets (line 2). Then, the negative template *types*, as well as the individual *constraints* in the *constraintSet* are defined (lines 3-4). In the following, we denote constraints of the types defined in line 3 as *negative constraints*. We then proceed to compute inconsistencies.

Trivial Inconsistencies (lines 6-9). We iterate over all negative constraints to verify if there is a corresponding complement template with the identical parameter signature. This is performed with the method *findRulesByTypeAndParams(type, params)*, which returns a set of sets. This function takes as input a specific template type and parameters (i.e. an ordered pair of tasks, e.g. *a,b*) and returns a set of sets, where the inner sets contain all corresponding constraints. As depicted in line 7, we pass the complement of the respective negative constraints as a parameter. Consequently, the function used in line 7 returns all trivial inconsistencies to the respective constraint *c*. A new minimal inconsistent subset containing a pair with the found constraint and constraint *c* is added to the set of minimal inconsistent subsets.

Generalization-Based Inconsistencies (lines 11-15). Section 2 defines the complements to the negative template types *NotCoExistence*, *NotSuccession* and *NotChainSuccession*. We denote these complements as the *complementSet* of the respective negative template. For all negative constraints γ, we iterate over all constraints of a template type contained within the *complementSet* of γ, with identical parameter signature (line 12). To compute inconsistencies, we again use the function *findRulesByTypeAndParams*. Here, we pass as parameters the current complement type to the current negative constraint, and the parameters of the negative constraint (line 13). This yields all generalization-based inconsistencies to the current negative constraint. We add a set containing the negative constraint and the computed contradicting constraint to the set of minimal inconsistent subsets (line 15).

Path-Based Inconsistencies (lines 17-21). After the previous two steps, all constraints in contradiction to *NotCoExistence*- and *NotChainSuccession* constraints have been identified. The detection of all contradictions to templates of type *NotSuccession* requires a path-based perspective, due to inconsistencies arising through transitivity. To find path-based inconsistencies, we construct a graph-like structure consisting of the relations between individual Declare constraints (line 17). We denote this as a task entailment graph.

Definition 2 (Task Entailment Graph). Let the declarative process model $\boldsymbol{M}$, the task entailment graph of $\boldsymbol{M}$ is a tuple $\boldsymbol{G} = (\boldsymbol{A}, \boldsymbol{E}, \boldsymbol{t})$, where $\boldsymbol{A}$ is the set of tasks in $\boldsymbol{M}$, $\boldsymbol{E} \subseteq \boldsymbol{A} \times \boldsymbol{A}$ is the set of directed edges between activities, and $\boldsymbol{t}$ is a function $\boldsymbol{t}: \boldsymbol{E} \rightarrow \boldsymbol{T}$, which maps an individual edge to a template type from $\boldsymbol{T}$ in $\boldsymbol{M}$.

An edge $\boldsymbol{e} \in \boldsymbol{E} = (a,b)$ indicates that the task *a* entails the task *b* via the constraint $\boldsymbol{t(e)}$. To compute path-based inconsistencies, we iterate over all negative constraints of type *NotSuccession* (line 18). Let the parameters of the respective negative constraint be *a* and *b*, then, we compute all paths from *a* to *b* in the task entailment graph (line 19). Each path from *a* to *b* constitutes a contradiction to *NotSuccesion(a,b)*, thus the path is added to the set of minimal inconsistent subsets (line 21).

We revisit the example constraint set from the example in Figure 3. All non-negative constraints are used to construct the task entailment graph, shown in Figure 8. The graph can be understood, such that task *b* is entailed by task *a* via the template type *ChainSuccession*. The rest of the graph is constructed accordingly.

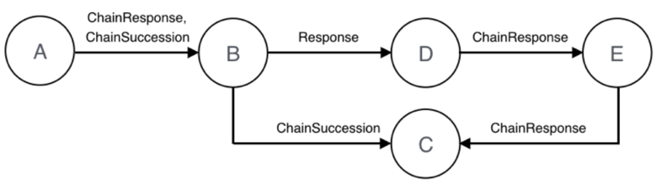
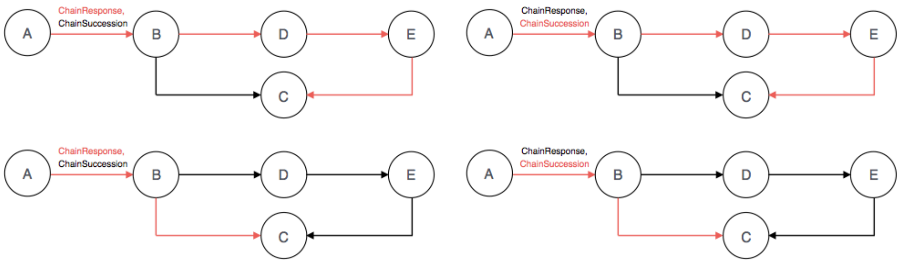


Figure 8. Task Entailment Graph for Example D_1 from Figure 3

The constraint set in this example also contained *NotSuccession(a,c)*. To search for a contradiction to this negative constraint, we verify if there are paths from *a* to *c*. In total, there are four possible paths from *a* to *c*, shown in Figure 9 (Note that the edge from task a to b in the graph has two types, due to the constraints *ChainResponse(a,b)* and *ChainSuccession(a,b)*). Via these four paths, *c* can transitively be entailed from *a.* Hence, all these paths contradict *NotSuccession(a,c)*. Accordingly, the constraint *NotSuccession(a,c)* and all constraints on the identified paths in the task entailment graph
respectively constitute minimal inconsistent subset.



1. **Figure 9.** Paths found from task *a* to task *c*, for Example D_1 from Figure 3

In our algorithm, following the computed minimal inconsistent subset, the constraint with the highest $C_\#$ culpability is deleted. In the example, this is *ChainSuccession(a,b)* with a $C_\#$ value of 4. The proposed algorithm would therefore delete this constraint to generate $d_{ChainSuccession(a,b)}(D_1)$. Continuing the example, the element with the subsequent highest culpability would be *ChainResponse(a,b)* = 3. A deletion would consequently eliminate all remaining ***MIS***, with a total information loss of two constraints. The sequence $\mathbf{\Gamma}_S$ traces the constraints which were deleted for further inspection, i.e. here $\mathbf{\Gamma}_S$ = <*ChainSuccession(a,b)* , *ChainResponse(a,b)*>.

To conclude, our approach computes all minimal inconsistent subsets with the above algorithm. Then, we delete the constraint with the highest culpability value w.r.t. the $C_\#$ measure. This process is iteratively repeated until the set of constraint is consistent.

4 Evaluation

We implemented our approach to resolve inconsistencies in Declare models in Java. Our evaluation is based on data from the Business Processing Intelligence Challenge 2017 (BPI). In the scope of the BPI, real life event logs are provided.

The analyzed log[1] of the 2017 challenge is an event log of a Dutch financial institute and comprises 262.200 events in 13.087 cases. From this log, we mined a declarative process model in the Declare language using Minerful, which is a state-of-the-art tool for declarative process discovery [6]. We mined three different process

[1] https://www.win.tue.nl/bpi/doku.php?id=2017:challenge

models, with the respective support factors of 75%, 85% and 95%. We then applied our algorithm to all three models. Table 1 summarizes our evaluation results.

Table 1. Results from the application of our algorithm to the BPI challenge log

Support Factor	75%	85%	95%
Discovered Constraints	305	232	207
Initial number of MIS	28954	731	639
Deleted Elements needed	5	1	1
Information Loss	1,63 %	0,43%	0,48%
Runtime	101099ms	9148ms	4695ms

The declarative model mined with a support factor of 75% (*M1*) consisted of 305 constraints. We computed over 28000 inconsistencies in this constraint set, which shows that the lowering the support factor can result in inconsistent models, even for state-of-the-art discovery algorithms. Our algorithm was able to resolve all these subsets in *M1* by deleting a total of only 5 constraints, which equals a total information loss of 1.6% of all initial constraints. This was possible by iteratively determining the constraint with the highest culpability, as deleting this constraint warrants the highest number of resolved ***MIS*** for each iteration. Figure 10 shows an overview of the number of ***MIS*** remaining after each iteration of our algorithm for *M1*. As can be seen, in each of the first three iterations, the number of MIS is reduced roughly by half. Then, a significant reduction from roughly 4000 to 700 MIS can be achieved in iteration 4. Figure 10 also shows the $C_{\#}$ culpability value of the respective constraint with highest culpability in each iteration. As can be seen, the first iteration resolves around 16000 of all 28954 MIS.

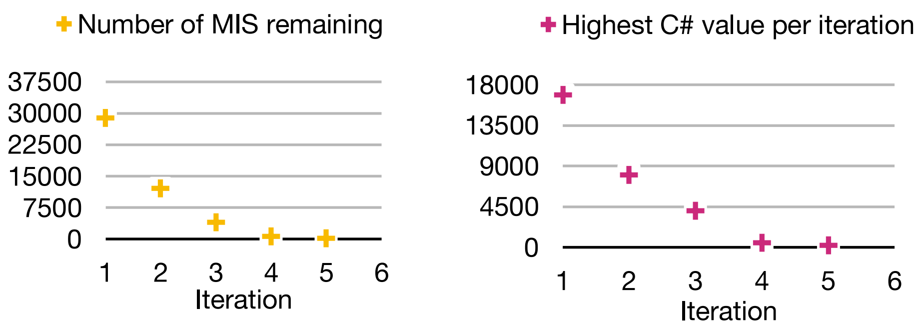


Figure 10. Number of MIS remaining (left) and highest C# culpability for a constraint (right) during the 5 algorithm iterations of resolving *M1*.

Specifically, for *M1*, the five constraints which needed to be deleted to completely resolve inconsistency are shown in Figure 11. The deletions as in Figure 11 can be presented to users. Should they deem this constraint must not be deleted under any circumstances, our approach can be extended with a „whitelist“.

Response(A_Incomplete, O_Accepted), $C_{\#}$ = 16890
Precedence(W_Validateapplication, W_PersonalLoancollection), $C_{\#}$ = 8020
Response(A_Incomplete, A_Pending), $C_{\#}$ = 3348
Precedence(O_CreateOffer, W_PersonalLoancollection), $C_{\#}$ = 496
Precedence(O_Created, W_PersonalLoancollection), $C_{\#}$ = 200

Figure 11. Constraints deleted corresponding to Figure 10.

For the other two process models, mined with 85%, resp. 95%, support factor, we observed an interesting case. The algorithm was able to make the set of constraints consistent by deleting only one element. In both cases, this was the constraint *NotCoExistence(W_PersonalLoancollection, O_Sent)*. Thus, this constraint was part of all ***MIS*** in both models. Apparently, the support factor configuration yielded exactly this one constraint which makes the entire model unusable. This is a good case for our approach idea of determining the constraint that has the highest blame in the overall inconsistency. Based on culpability measurement, our algorithm could effectively resolve all inconsistencies in a low runtime, deleting only one constraint.

Our results show, that even for a low number of constraints, the number of inconsistencies can be rather large. This underlines the need for post-processing techniques in process discovery such as our approach. Our algorithm was able to resolve these inconsistencies.

5 Related Work

This work is related to consistency checking in Business Process Management (BPM) and declarative process discovery.

Consistency checking in BPM is widely recognized as a challenging task [2-6, 13, 15]. A core task here is to ensure the consistency and correctness of BPM related artifacts. We have focused on the area of verifying the consistency of declarative process model artifacts, respectively a resolution of inconsistency. There have been many proposals for the *qualitative* analysis of the reasons of inconsistencies [1, 7, 14]. However, works such as [13, 15, 19] point out the benefits of a *quantitative* analysis. The intuition here is that individual problems can have a different severity. Hence, using such a quantitative assessment as proposed offers a more sophisticated insight on how to resolve inconsistencies in declarative constraint sets [5, 13].

As a directly related work, we identify the report by Di Ciccio et al (2017). Those authors also proposed an approach to resolve inconsistencies in Declare constraint sets. Those authors however do not consider all ***MIS***, but rather build a maximal consistent *new* constraint set. As discussed, we adapt the approach by [9] of resolution by *deletion*. The authors in [6] point out, that the computation of inconsistencies by comparing all possible subsets of constraints is intractable. We agree, that this would be indeed unfeasible. To solve this problem, we therefore defined different inconsistency types in Declare, based on the set of pre-defined template types, which allows for informed comparisons and feasible computation. Our evaluation shows, that the computation of ***MIS*** can be performed in a feasible run-time.

Our algorithm promotes declarative process discovery. Due to the introduced notion of a support factor, existing process discovery techniques can yield inconsistent declarative process models [6, 17]. Our work contributes with an approach for automated resolution, allowing for effective post-processing techniques in process discovery.

6 Conclusion

In this work, we presented an approach for resolving inconsistencies in declarative process models. Our approach was implemented for process models in the Declare language. Our evaluation based on a real-life event log showed that it was possible to compute and resolve around 28.000 minimal inconsistent subsets in a feasible runtime. In the process model mined with 75% support factor in our evaluation, there were 305 constraints. However, only 87 constraints were part of any ***MIS*** (i.e. all other constraints had a $C_{\#}$ value of 0). We therefore argue it makes sense to only consider those constraints that contribute towards inconsistency, as in our approach. Deleting constraints with a $C_{\#}$ value of 0 cannot decrease inconsistency in Declare [9]. Deleting the element with the highest value resolves the most ***MIS***. To the best of our knowledge, this work is the first to incorporate culpability measurement.

In our evaluation, all models could be resolved with a deletion of max. 5 constraints, which equaled an information loss of under 2% in all models. This shows that our algorithm is aligned with our goal of ensuring a low amount of information loss. As a direct limitation, elements to delete are automatedly selected. Hence, it would be possible that knowledge which domain experts would not delete is automatedly deleted. To solve this problem, we store the sequence of deleted constraints for further inspection by domain experts. The aim of this paper was to present an approach for fully automated resolution approach. In future work, we will include the possibility to define constraints which must not be deleted. In case such „marked“ constraint would have the highest culpability value in an iteration run, it would simply be possible to delete the constraint with the next highest $C_{\#}$ value based on the computed culpability ranking.

A limitation of this work is, that our algorithm considers only the predefined set of Declare templates. While the majority of declarative discovery tools only focus on these templates, future work should be directed to find ***MIS*** in arbitrary constraints.

As a key learning, we observed that the support factor used during declarative process mining strongly impacts the resulting model. This should be considered by companies seeking to implement declarative process discovery. Automated approaches as presented in this work should be utilized to verify the consistency of the resulting models to allow for correct and compliant process execution.

References

1. Awad, A., Smirnov, S., and Weske, M. (2009). Towards resolving compliance violations in business process models. GRCIS. CEUR-WS. org.
2. Batoulis, K., Nesterenko, A., Repitsch, G., and Weske, M. (2017). Decision management in the insurance industry: Standards and tools. In Proceedings of the BPM 2017 Industry Track (BPM 2017), Barcelona, Spain, September 10-15, 2017., pages 52–63.
3. Batoulis, K. and Weske, M. (2017a). Soundness of decision-aware business processes. In International Conference on Business Process Management, pages 106–124. Springer.
4. Calvanese, D., Dumas, M., Laurson, U., Maggi, F. M., Montali, M., and Teinemaa, I. (2016). Semantics and analysis of dmn decision tables. In International Conference on Business Process Management, pages 217–233. Springer.
5. Corea, C. and Delfmann, P. (2018). Supporting business rule management with inconsistency analysis. In Proceedings of the BPM 2018 Industry Track (BPM 2018), Sydney, Australia, September 09-14, 2018.
6. Di Ciccio, C., Maggi, F. M., Montali, M., and Mendling, J. (2017). Resolving inconsistencies and redundancies in declarative process models. *Information Systems*, 64:425–446.
7. Ghose, A. and Koliadis, G. (2007). Auditing business process compliance. In International Conference on Service-Oriented Computing, pages 169–180. Springer.
8. Graham, I. (2007). Business rules management and service oriented architecture: a pattern language. John wiley & sons.
9. Grant, J., & Hunter, A. (2011). Measuring consistency gain and information loss in stepwise inconsistency resolution. In European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (pp. 362-373). Springer, Berlin, Heidelberg.
10. Grant, J. and Martinez, M. V. (2018). Measuring Inconsistency in Information. College Publications.
11. Hunter, A., Konieczny, S., et al. (2008). Measuring inconsistency through minimal inconsistent sets. *KR*, 8:358–366.
12. Imgrund, F., Malorny, M., and Janiesch, C. (2017). Eine Literaturanalyse zur integration von business rules und business process management. 13. Internationale Tagung Wirtschaftsinformatik, WI 2017, St.Gallen, 2017.
13. Lu, R., Sadiq, S., and Governatori, G. (2008). Measurement of compliance distance in business processes. Information Systems Management, 25(4):344– 355.
14. Polyvyanyy, A., Ouyang, C., Barros, A., and van der Aalst, W. M. (2017). Process querying: Enabling business intelligence through query-based process analytics. *Decision Support Systems* 100: 41-56.
15. Sadiq, S. and Governatori, G. (2015). Managing regulatory compliance in business processes. In Handbook on Business Process Management 2, pages 265–288. Springer.
16. Thimm, M. (2016). On the compliance of rationality postulates for inconsistency measures: A more or less complete picture. *Künstliche Intelligenz*.
17. van Der Aalst, W. M., Pesic, M., and Schonenberg, H. (2009). Declarative workflows: Balancing between flexibility and support. Computer Science-R&D, 23(2):99–113.
18. van der Aalst, W. M. P., La Rosa, M., and Santoro, F. M. (2016). Business process management. Business & Information Systems Engineering, 58(1):1–6.
19. Weidlich, M., Mendling, J., and Weske, M. (2011). Efficient consistency measurement based on behavioral profiles of process models. IEEE Transactions, 37(3):410–429.
20. Weske, M. (2010). Business process management: concepts, languages, architectures. Springer Publishing Company, Incorporated.

Strategic Analysis in the Realm of Enterprise Modeling – On the Example of Blockchain-Based Initiatives for the Electricity Sector

Sybren de Kinderen[1], Monika Kaczmarek-Heß[1], Ivan Razo-Zapata[2], and Qin Ma[3]

[1] University of Duisburg-Essen, Essen, Germany
{sybren.dekinderen,monika.kaczmarek}@uni-due.de
[2] National College of Ireland Dublin, Ireland
ivan@coca-ci.org
[3] University of Luxembourg, Esch-sur-Alzette, Luxembourg
qin.ma@uni.lu

Abstract. This paper continues our previous work on modeling support for strategic analysis, by (1) extending a proposed modeling language for strategic analysis (called SAML), among others, with features proposed by business scholars to increase the expressiveness of the analysis, and (2) relating the extended SAML to a language for IT infrastructure analysis (called ITML). Thus, we explicitly contextualize strategic analyses by accounting for the role and impact of IT infrastructure. A scenario in the electricity industry is used to illustrate the analysis proposed.

Keywords: strategic analysis, conceptual modeling, NRGcoin.

1 Introduction

Strategic management, to which strategic analysis belong, can be considered as "a collection of decisions and actions taken by the business management in consultation with all levels within the company to determine the long-term activities of the company" [1, p. 125]. The aim of the actions to be undertaken may be manifold, among others, improvement of the competitive position, or the realization of profit growth. In the era of digital transformation those improvements are often achieved through the application of IT artifacts [2].

In this paper, we focus on a specific area of strategic management, namely strategic formulation [3, 4] and investigate strategic analysis tools and approaches, which can be used to support assessment of, among others, planned (digital) initiatives. In this context, a SWOT analysis (Strength, Weakness, Opportunity and Threat Analysis) is an instrument that is traditionally used [5]. However, although well-established and often used, SWOT is also considered to be vague and oversimplified [5]. As a response, approaches that extend SWOT have been proposed, which, among others, (1) suggest additional organizational aspects on which a SWOT analysis can be conducted [6], such as organizational culture and technologies, or (2) propose to combine a SWOT analysis

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

with the resource-based view [7]. The conceptual modeling community have also acknowledged the need for better instruments supporting strategic analysis [8], and proposed different approaches, cf. [9, 10]. However, they either do not provide semantically rich concepts, are still in their development phase, and/or fall short when it comes to integration with other perspectives (particularly the IT perspective).

Motivated, on the one hand, by this gap, and on the other hand, by the increasing number of digitalization initiatives, in this paper we introduce an instrument for a model-based strategic analysis that explicitly accounts for IT infrastructures. We follow the design school of strategy, cf. [4], thus, we focus on establishing the fit between the internal capabilities and external possibilities [11]. Since the goal of the proposed modeling method is to allow to rationalize the decision made, we treat IT as a white box and consider the large range of both its internal and external factors. To show the applicability of the proposed approach, we focus on the smart grid domain, being one of the domains heavily affected by digital transformations [12]. Particularly, we focus on the blockchain-based NRGcoin initiative in the energy sector [13].

In this paper, we continue and extend our earlier work in the area of multi-perspective valuation supported by modeling, cf. [14], and deliver a three-fold contribution: (1) we extend a Strategic Analysis Modeling Language (SAML) with additional features as found through a literature analysis. Prominently, the proposed extensions include accounting for additional features as proposed by business scholars, and comparing alternative strategic elements, partly on the basis of a root-cause analysis, partly on the basis of relations to other modeling languages; related to the latter, (2) we introduce a mapping between SAML and a language for expressing IT infrastructures called ITML [44]. Thus, we deepen the relation between two particular modeling languages, compared to the brief description of the different relations within the landscape of six languages in [14]; and (3) benefiting from the performed extensions, we deepen a strategic analysis of the NRGcoin initiative. To extend the relationships between SAML and ITML, and to extend SAML, we follow the method proposed by [15] (see Section 3).

The paper is structured as follows. First a short introduction to the strategic analysis with the main focus assigned to a SWOT analysis, as well as the role of conceptual modeling is provided. Then, we describe the extended SAML and its connections to ITML. Next, the case study is discussed showing the applicability of the proposed approach. The paper concludes with final remarks.

2 Background

Strategic Analysis: A strategic perspective emphasizes the long-term outlook on an organization or a network of organizations. This long-term outlook informs analysis of an initiative to be undertaken, e.g., in terms of the long-term organizational goals being pursued and the influence these have on the value-exchanges taking place [16]. Here, strategic orientation refers to analyzing, for a particular organization, the fit between its external situation and internal characteristics [17]. Such analyses are typically done with traditional business school instruments, prominent ones being the 5 Forces approach or Value Chain [18], balanced scorecard [19], and SWOT analysis [5]. For the remainder of this paper we focus on SWOT, since (1) SWOT, in part due to its simplicity, is still an often-used approach to support the strategy formulation [5], and (2) its shortcomings – being conceptually vague, a missing relation to organizational aspects other than strategy – make extensions to SWOT necessary.

SWOT [20] can be used to assess qualities internal to an organization in terms of Strengths (S) and Weaknesses (W), and situations external to the organization in terms Opportunities (O) and Threats (T). Furthermore, SWOT allows one to compare internal qualities to external situation, thus allowing for a so-called strategic fit analysis [5]. A typical SWOT analysis lists favorable and unfavorable internal and external issues in the four quadrants of an analysis table, thus providing a better understanding "how strengths can be leveraged to realize new opportunities and understand how weaknesses can slow progress or magnify organizational threats" [5, p. 1]. Although SWOT is a popular tool, it is considered to be **vague** and **oversimplified** [5], as it constitutes merely a list that does not "provide sufficient context for adequate strategy optimization" [5]. In line with these criticisms, authors of [8] state that (1) concepts of strategic analysis approaches are often ill-defined, with "strength" having a colloquial understanding at best, which is especially lamentable given the contingency of this concept [8, p. 47], and that (2) a relation of the strategic approaches to a detailed understanding of other aspects of an organization is often missing, with, e.g., IT infrastructure being treated as a black box [8, p. 48].

To address this oversimplification, business scholar literature combines SWOT analyses with other approaches. For one, [6] adopts the resource-based view (RBV) of the firm to provide a further assessment of the strengths and weaknesses internal to an organization, as identified per SWOT, so as to make a comparison with competitors. For instance, for a given "strength" one can use the RBV to assess its rarity, substitutability, and how easily the strength can be replicated (or imitated) by others. Yet another example is a telescopic observations strategic framework [7], which maps strengths, weaknesses, opportunities, and threats against suggested categories, such as technological advancements, economic considerations, legal, and regulatory requirements. Furthermore, to prioritize SWOT items, SWOT has been extended with (quantitative) methods, among others, Analytic Hierarchy Process (AHP)-SWOT [21], Analytic Network Process (ANP)-SWOT [22] and Importance-Performance Analysis [23].

Conceptual modeling in support of strategic analysis: To address ill-defined concepts, and a lacking relation to other aspects of an organization, conceptual modeling in general, and enterprise modeling in particular, can play an important role. Indeed, modeling techniques exist that support strategic analysis, in terms of, among others, goal-oriented requirements engineering (GORE), and modeling techniques that explicitly incorporate concepts from business scholar literature on strategic analysis. Regarding GORE, there exist a variety of conceptual modeling techniques, such as i* [24], the Goal-oriented Requirements Language (GRL) [25], or GoalML [26], for a recent overview cf. [27]. With their focus on modeling (short/medium/long)-term goals, these techniques form a useful point of departure for strategic analysis and have also been used to that extent, cf. [28]. However, the focus on goals means that these general

GORE techniques – with a few exceptions – are equally applicable to other types of analyses. Thus, in their key concepts they often do not address ideas pertaining to strategic analysis as we find them in the discussed business scholar literature.

Besides these generic techniques, however, there exist also GORE approaches that do explicitly take on board strategic analysis concerns from business literature. In line with the idea of analyzing the strategic fit, i* explicitly recommends to make both an analysis of the goals internal to actors, and an analysis of the interactions between actors [24]. Likewise, in [29] the authors provide an approach to analyze strategic fit by combining a domain-specific modeling language (which includes concepts such as goals, value proposition, activity, process, competence, etc.) with AHP and heat mapping techniques. The Business Intelligence Model (BIM) [9] offers concepts (e.g., goals, situations, influences, and indicators) to support strategic business analysis in terms of, both (1) continuous monitoring of organizational goal fulfillment based on KPIs, and (2) analyzing the strategic fit, particularly in terms of a model-based SWOT analysis [9]. The defined relations allow to reason on relationships between situations, influences, and indicators. Although the BIM approach seems to be a powerful tool, to the best of our knowledge, it is not integrated with other perspectives, i.e., with elements from the action system and the information system of an enterprise. As a consequence, BIM does not allow for more sophisticated analyses and strategy definition. Finally, some interesting initiatives may be observed to model strategic plans or strategic control in the realm of enterprise modeling, e.g., [10, 30], which aim at capturing, e.g., influence or impact of an initiative on an enterprise and its resources. However, those initiatives either are still in the development phase, or do not consider IT as a first-class citizen.

Existing enterprise modeling (EM) approaches, while providing the possibility to express various views on an organization, do not explicitly account for strategic analysis as defined by business scholars, cf. [14]. Yet, approaches exist that focus on the (strategic) analysis of IT infrastructure. For instance, ArchiMate [31] allows for relating IT infrastructure and strategy, in the sense that: (1) its motivation extension [31, p. 80] allows for expressing strategy concepts, as also reflected in the explicit mappings between ArchiMate's general motivational concepts (e.g., "Goal") and concepts from the business scholar discourse on strategy (e.g., "Mission"), cf. [32]; (2) ArchiMate provides a rudimentary expression of IT infrastructure elements; and (3) ArchiMate provides the ability to express various relationships between layers [31, p. 107]. However, being a language to express enterprise architecture concepts in general, ArchiMate's focus is not a strategic analysis of IT infrastructure per se. Therefore, various extensions have been proposed, e.g., [33, 34]. For instance, ArchiMate has been extended to relate business goals to IT projects and their underlying infrastructure [34]. The aim of this was to enable valuation of IT portfolios using ArchiMate together with Bedell's method, in order to measure the strategic importance of IT infrastructure to organizations' goals. Nonetheless, the overall method only focuses on analyzing IT portfolios of a single organization. Moreover, while concepts from both IT infrastructure and strategy play a notable role here, the focus is actually placed on quantitative valuations. As such, the particular characteristics of IT infrastructure, and the implication these have for the organizational strategy, are less of a focus. Similar to ArchiMate, ARIS [35] offers concepts to analyze an organization from both a strategic perspective and an IT infrastructure perspective, which provides the possibility to relate these perspectives. However, while ARIS offers more expressiveness than ArchiMate, as especially visible on the strategic end which combines ARIS's concepts with a balanced scorecard analysis [35, p. 187], the specific relation between IT and strategic analysis remains under-explored. Finally, in proposing a method for model-driven business-ICT alignment, [16] relates organizations' strategic perspectives (described in e3forces) to their corresponding IT/IS perspectives. However, the model-driven strategic analysis pertains mostly to the external market level only, whereas we require a focus also on the strategic analysis of *internal* (IT) resources. Also, the IT infrastructure is mostly depicted in an informal (arrow-and-boxes like) manner, which inhibits its differentiated analysis.

Strategic analysis of smart grid (SG) initiatives: The emergence of smart grids is driven by the convergence of information and power delivery technologies [12]. The cornerstone of a smart grid is the ability for intelligent devices (e.g., smart meters), dedicated software, processes, etc., to interact and cooperate via an ICT infrastructure. SG initiatives, as any digitalization initiative, are analyzed using the already mentioned instruments, e.g., by means of a SWOT analysis [36]. There are, nonetheless, domain-specific initiatives that aim to evaluate digitalization initiatives in the SG sector, too. For instance, the Smart Grid Maturity Model (SGMM) assesses smart grid initiatives by focusing on (1) six maturity levels (Level 0 Default to Level 5 Pioneering), and (2) eight domains (logical groupings of SG related characteristics). In a similar vein, [37] provides a method to assess the strategic value of IT in SG initiatives by combining the Smart Grid Architecture Model (SGAM), an enterprise-wide and service-oriented framework to describe SG architectures [38], and the Bedell method, which computes the effectiveness and importance of IT elements [39, 37].

3 Modeling Support for Strategic Analysis of IT

Based on the conducted study of business scholar literature and existing work in conceptual modeling, a set of requirements has been identified. Due to space restrictions, we present them clustered into three main postulates.

Postulate 1: Integrating strategic analysis concepts with other elements of an enterprise action system and an information system (IS). *Rationale*: [40] suggests that one of the primary misuse of the SWOT tool is not to link it with the other perspectives on an organization. Indeed, to make a rational decision and decide on the strategy to follow, considering aspects of an enterprise action system (business processes, goals, resources etc.) and IS, is important [4, 40]. In the era of digital transformation, especially the latter, i.e., accounting for the IT perspective, becomes crucial [2]. Indeed, if a modeling language is explicitly related to concepts expressing the IT perspective, one may conduct a strategic analysis that is grounded in the actual IT capabilities of an organization.

Postulate 2: Provision of well-specified, semantically rich concepts which account for proposed extensions to SWOT-based analysis. *Rationale*: Considering two mostly criticized aspects, namely vagueness and oversimplification of concepts used during the analysis process (cf. Section 2), a modeling method supporting a strategic analysis should provide a rich set of domain-specific concepts with a rich set of attributes that one could use during the analysis process. Here, the proposed extensions to SWOT, such as, e.g., the telescopic observations strategic framework [7] could be accounted for. It should be also possible to assign different weights and different probabilities to different situations in order to mark their importance, cf. [23], as well as to input a justification for the assigned classification (e.g., why we consider something as a strength). When it comes to the classification of a given situation, the modeling approach should enable the classification of some situation in some context differently to account for the fact that "external (or internal) factors of an organization are not always opportunity (strength) or threat (weakness); in other words, in different conditions, they have different meanings" [41].

Postulate 3: Accounting for a rich set of relationships, as proposed in the literature. *Rationale*: a SWOT analysis leads to the creation of a table of SWOT items, cf. [5]. As such, it does not account for the complexity of the phenomena and resulting consequences. Indeed, if instead of a table we use a diagram, we are able to represent a network of concepts connected using different relationships. This allows to conduct a more sophisticated analysis. Among others, SWOT is often criticized for providing "no indication of causality among the strengths and weaknesses, nor are they ranked into any hierarchy" [42, p. 5677]. As a response, the said set of relationships should allow to account for causality relations among all SWOT concepts (e.g., cause, effect) as well as to account for hierarchies of different states.

Language design: As we are interested in the integrated view on an enterprise, making a strategic analysis modeling language part of one of the existing enterprise modeling approaches seems to be reasonable. Based on the postulates, we need a language architecture that would, among others, (1) support the definition of semantically rich concepts, i.e., it should allow for expressing attributes and constraints; and (2) account for various aspects of organizational action system and information system. The MEMO family of languages [43], which we already extended in our previous work, cf. [14], has a language architecture fulfilling the stated postulates[1], and accounts for various perspectives on an organization. Therefore, we continue our previous work with MEMO and use the MEMO Meta Modeling Language (MML) [15] to make necessary extensions.

Fig. 1 (the upper part) shows the key concepts of the extended Strategic Analysis Modeling Language (SAML) [14] and their connections to concepts from other MEMO languages such as GoalML (goal modeling) or OrgML (organization structure modeling) [43]. The initial version of SAML has been extended with additional concepts, properties, relationships and constraints, as indicated in Fig. 1. In terms of the employed language design method [15], it is notable that (1) we consider the purposes and use scenarios as first-class citizens that drive the design of the language landscape, (2) we employ the guidelines for concept inclusion from [15]. For example, the concept *involvementContext* (see Fig. 1) and its various attributes and relations conform to both

[1] For an elaborate discussion on the selection of the approach to be extended, see [14].

the guideline "relevance", in terms of relevance to various analysis scenarios, and "invariant semantics", in the sense of its semantics being invariant over different analysis scenarios, as well as the concept having its own essential characteristics. Note that, due to space constraints, we unfortunately cannot further elaborate on the used language design method.

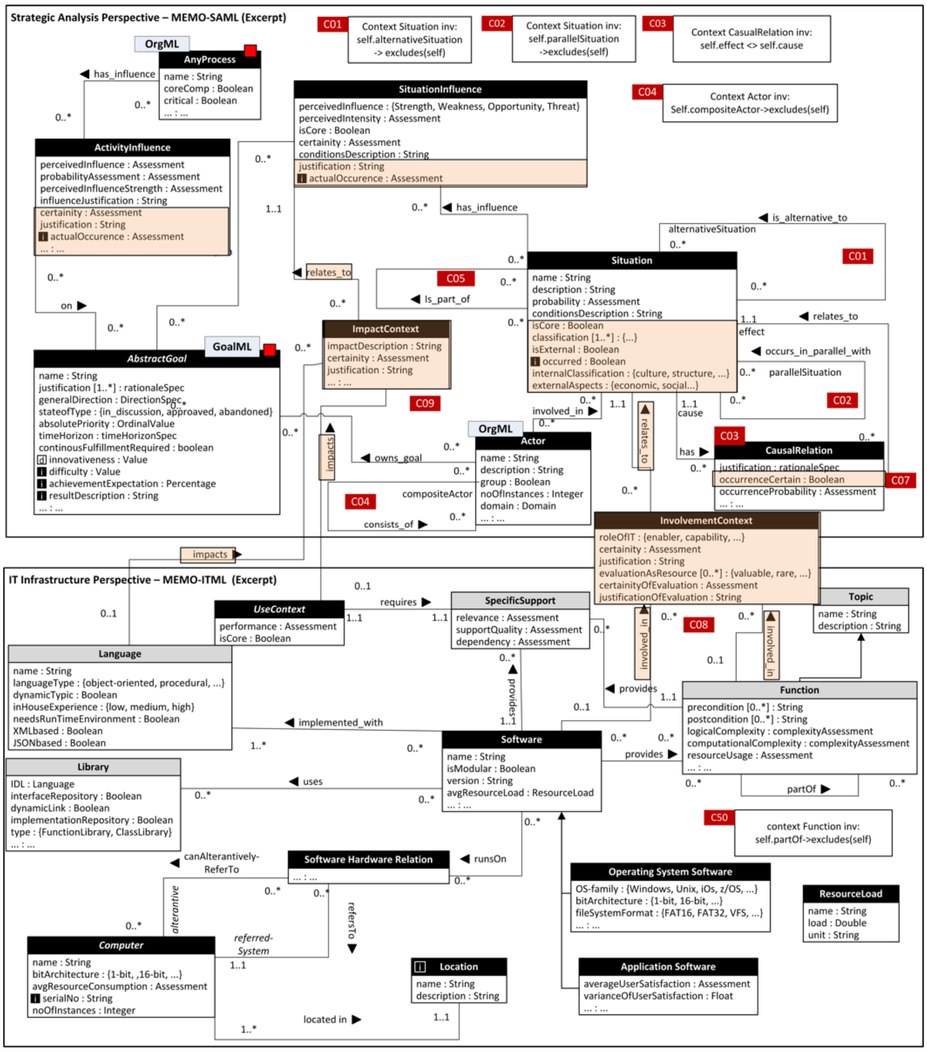


Figure 1. SAML and exemplary relations to the extended ITML

The main concept for analysis is a *Situation*, which, in line with BIM, is defined as a partial state of the world that has a structure consisting of relations and elements. In line with the presented postulates, a *Situation* has a rich set of attributes allowing to describe which state we mean, what is its probability, as well as, e.g., its classification (e.g., whether it is internal or external [4, 1]). For all concepts, in line with Postulate 2, the level of justification, scope, or importance can be defined. *Situations* may be linked to other *Situations* by relationships: (1) *is_alternative_to*, which allows us to model SWOTs arising from different alternatives in the same diagram. This is opposed to a typical SWOT analysis, whereby one has to draw a separate table for each alternative, which also makes it difficult to compare individual SWOT elements to each other; and (2) *occurs_in_parallel_with*, which allows us to cluster situations according to a logical grouping; and (3) *CausalRelation*, which allows to account for the fact that situations are not independent, cf. [42]. Here, to streamline the analysis, we differentiate in the concrete syntax between two types of relationships considering the probability of occurrence (cf. attribute *occurrenceCertain*): *results_in* and *may_lead_to*. Next, to aggregate situations into a more abstract situation, situations can be modeled as a hierarchy using the relation *is_part_of*. Finally, to account for the fact that depending on the context the same situation can be differently classified, we introduce *SituationInfluence* as an Association Class with a set of relevant attributes.

Furthermore, we include explicit relationships to elements of an IT infrastructure (i.e., to ITML [44]), such as *impacts*, *involved_in*, which allow us to identify the role and the type of influence of the IT infrastructure on possible situations. To increase the semantics of those relationships, we benefit from the already mentioned resource-based-view and characterize those relationships with an additional set of attributes such as: (1) evaluation of the IT as resource as valuable, rare, imperfectly imitable, or non-substitutable; as well as (2) assessing the certainty of an evaluation and its justification. In addition, it is possible to assign a role to an IT artifact in a given situation. Here, we differentiate the role by referring to classifications that point out the role of information systems in innovation processes (e.g., as an enabler or capability, cf. [2]).

To ensure that the created models are consistent with the underlying language specification, we extend the corresponding tool for modeling with MEMO, called MEMO4ADO [45]. Finally, MEMO4ADO is used to model the NRGcoin scenario.

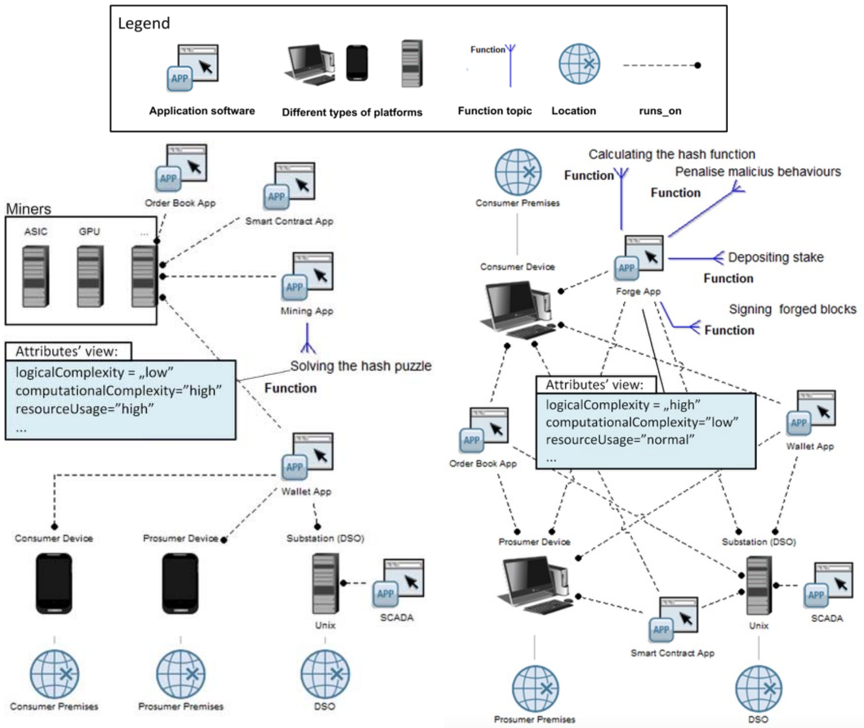
4 Illustration: Strategic Analysis of the NRGcoin Initiative

NRGcoin defines a blockchain-based support policy for renewable energy sources that aims to reward production and consumption of renewable energy [13]. The locally produced electricity is directly fed into the grid and withdrawn by consumers and prosumers. The electricity injected by prosumers is rewarded with NRGcoins, based on the amount of demand within a district the injection helps match. If the injected electricity does not match any demand, it is not rewarded, which encourages prosumers to consume their own electricity [13]. NRGcoins are rewarded to prosumers every 15 minutes, once smart meters inform the distribution system operator (DSO) about the amount of electricity being injected and withdrawn [13]. Thus, during the execution of the NRGcoin initiative, three types of transactions take place: (1) electricity consumption and injection transactions that record the amount of electricity being withdrawn/injected; (2) NRGcoin payment transactions that record: the number of NRGcoins being paid to the DSO for using the electricity grid infrastructure, the number of NRGcoins rewarded to prosumers for their injections, and the number of NRGcoins charged to consumers for their consumption; (3) NRGcoins trading transactions that record how NRGcoins are exchanged in the coin market against fiat currencies.

A tamper-proof ledger, as the one offered by a blockchain, serves as a promising mechanism to record these three types of transactions. Blockchain ledgers are distributed among the nodes that participate in the network. Transactions are organized in blocks that are chained up into blockchains. Every node in the network has access to the whole history of the transactions, and can check the validity of the blocks and transactions. Lacking a central authority to keep track of, validate, and write new records in such a distributed ledger, blockchain-based solutions make use of "consensus mechanisms" to reach agreement among nodes on who will (1) validate the transactions, (2) create the next block, and (3) broadcast it to the rest of the network [46].

In our previous work [14], we have demonstrated the advantages of NRGcoin initiative in achieving the following goals of stakeholders: (1) the share of green energy consumption is increased as consumers can purchase green energy at a fixed rate of NRGcoin; (2) self-consumption is promoted for prosumers because injection that does not match local demand will not be rewarded; (3) stress on DSO grids is relieved because local demand is met by local supply, and extra supply is self-consumed by prosumers, hence, there is less energy that needs to be transferred further up to the grid; (4) utilities' operational costs are reduced as most of the daily operations are automated with the help of smart contracts; and last but not least, (5) no dedicated budget from the government is needed because incentives to both green energy consumption and green energy production come from NRGcoin itself. In this paper, we address the problem of deciding which consensus protocol should be used to create the blocks in the NRGcoin ledger: Proof of Work (PoW) or Proof of Stake (PoS), cf. Fig. 2.

PoW is a pure cryptography consensus mechanism whereby so-called miners interested in becoming the creator of the next block compete to solve a cryptographic puzzle [46]. The first miner that finds the correct solution to the puzzle will become the creator of the next block. This miner will be rewarded with new coins (referred to as mined coins) and earns also fees associated with the validated transactions. In turn in PoS, blocks are said to be "forged" or "minted" instead of "mined". Candidates for the creator of the next block are referred to as validators. The probability for a validator to become actually the creator of the next block is in proportion to the amount of coins the validator owns. The selected validator earns fees associated with the validated transactions. Since the PoS technique is prone to security issues due to its simplicity [47, 48], various socio-economic counter-measures have been introduced. E.g., validators can be requested to lock a certain amount of coins as a stake in a security deposit in order to become a candidate for the next block. If the selected validator conducts malicious behavior while validating the next block, this validator will be punished by losing the stake (economical measure). A similar punishment measure can also be put in place from the social point of view, whereby the selected validator is required to sign the block she/he creates. If a peer node detects faults in the block and reports it to the network, the validator will be punished with bad reputation and will be forbidden to participate in future validations. In an extreme case, the node can even be expelled from the network.



(a) PoW IT infrastructure overview (b) PoS IT infrastructure overview

Figure 2. Overview on IT Infrastructure of Alternative Consensus Mechanisms

The decision on the used consensus protocol (PoS versus PoW) is important, as the protocols exhibit differing characteristics, which directly or indirectly impact the defined goals. We elicit and compare these characteristics by means of a SAML diagram for strategic analysis in combination with an ITML diagram for IT infrastructure analysis. Note that we focus only on those parts of the analysis that illustrate the added value of our approach, i.e.: (1) comparing alternative situations; (2) allowing to analyze how situations form a SWOT for achieving goals; and (3) enabling analysis on how situations relate to each other in terms of a root-cause analysis.

First of all, being a renewable energy initiative itself, NRGcoin's main goal is to promote energy efficient solutions (G1) as well as to promote production and consumption of renewable energy (G2). NRGcoin users are consumers and prosumers at the lower end of the electricity grid with the aim to set up a local energy community among them (G3) that calls for social responsibilities (G4). Achievement of the aforementioned goals can be examined from various perspectives. In the following, we show how the respective IT infrastructures underlying the PoW and PoS protocols (depicted in Fig. 2a and Fig. 2b respectively) contribute to the fulfillment of the goals of the NRGcoin initiative.

Looking at Fig. 3, we notice that the distinct features of PoW and PoS lead to two alternative *Situations* being the roots for two chains of Situations connected via causal relations differentiated into “results in” or “may lead to”, as explained in the previous section. Looking at the left side, among others, the mining application of PoW is simple, but very heavyweight, as the simple logic of PoW needs to be repeated several times until a hash code satisfying the condition can be found. As a consequence, mining takes on average long to finish, hence it likely causes delay to the validation of transactions (G5). In contrast, concerning time consumption, we find that PoS provides timely validation. In the diagram this is indicated by the situation “timely validation, little to no delays”, which subsequently, among others, traces back to a function topic “sign forged blocks” (an IT infrastructure element). Importantly, this tracing is done via an annotated SAML-ITML relation that (in line with the resource-based view of the firm, cf. the meta model in Section 3) is, both (1) “valuable”, with the justification that for PoS the signing of forged blocks happens in a short time compared to mining in PoW, and (2) “rare”, with PoS being novel and thus, not yet widely adopted. Moreover, due to its heavy weight, mining consumes a significant amount of electricity, which makes PoW a less energy efficient solution (G1). In our models, we first trace this significant electricity consumption to the situation “increased consumption of energy”, which has a relation “Weakness: very high” with goal G1. Subsequently, we trace this situation to the function topic “Solve the hash puzzle”, whose specific properties and underlying IT infrastructure we can further examine with the help of the corresponding IT infrastructure diagram (Fig. 2). More specifically, in the IT infrastructure diagram we find that the function topic “Solving the hash puzzle” is provided as a functionality of a “Mining App”, and that this functionTopic has the attribute “resourceUsage=high”. Furthermore, the resource use is also reflected in the underlying hardware required to run the mining application which often runs on dedicated, resource-intensive hardware (cf. Fig. 2a).

Considering that the validation is a time-consuming task, the management of the NRGcoin ledger may be outsourced to a set of miners, as indicated by the relation “may lead to”. This situation has two implications. Firstly, miners, being outside of the local energy community, have no access to the green energy produced within the community, hence would rather sell all earned NRGcoins than use them to purchase green energy for consumption. Therefore, this situation (being external from the point of view of our initiative) is a threat of PoW with respect to “G2: Promote Production and Consumption of Renewable Energy”. Secondly, miners have no sense of belonging to the community, hence their participation in validating transactions and mining is purely profit-driven. This is a threat towards “G4: Increasing Social Responsibility”.

In contrast to PoW, the PoS algorithm is complex (as it needs to integrate additional functions that implement social and financial counter-measures, Fig. 2b), but lightweight. As PoS is lightweight, validators consume modest amount of electricity to fulfill their tasks, hence PoS is an energy efficient solution (G1). Therefore, common

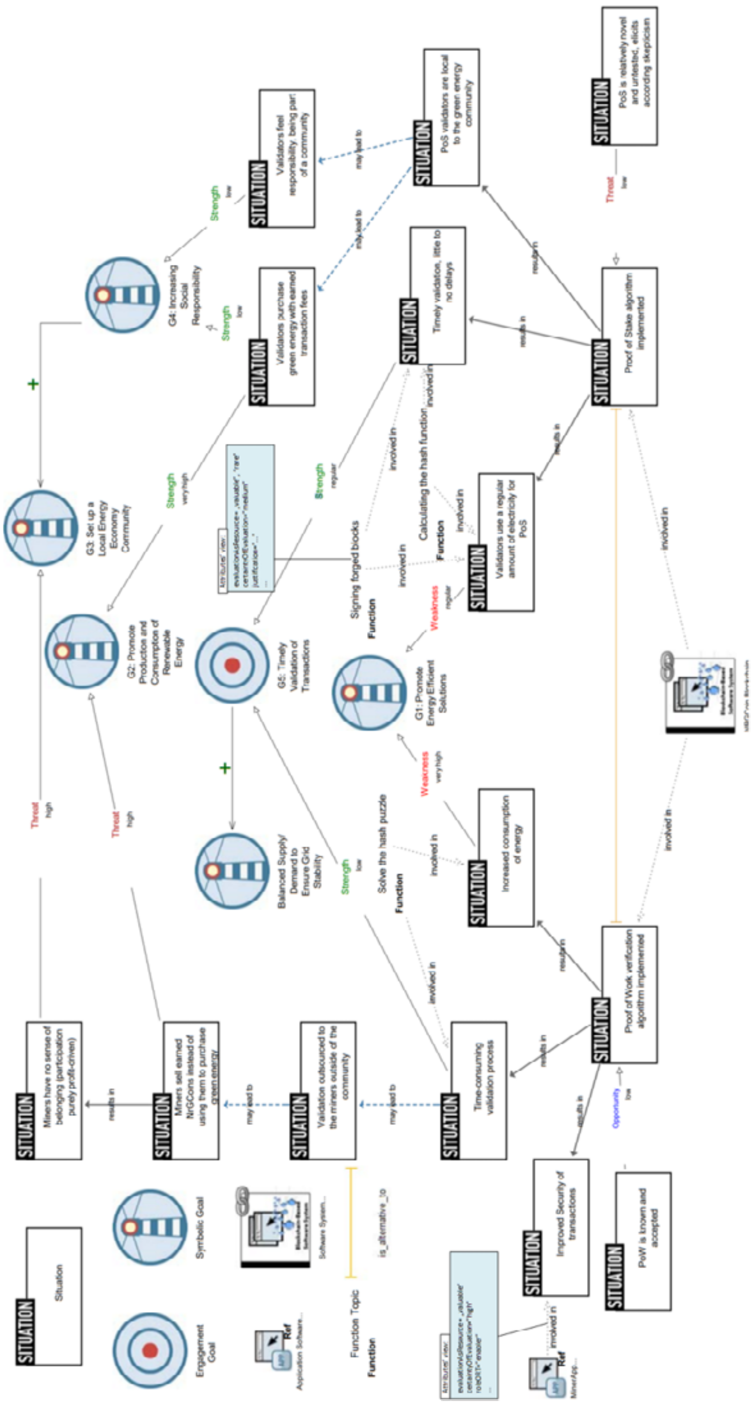


Figure 3. An excerpt of the Strategic Analysis Diagram of NRGcoin: PoW vs. PoS

PCs, which almost every family is in possession of, suffice to run the light PoS algorithm efficiently. This situation is a strength of PoS, because it allows any member of the NRGcoin community to become a validator. Unlike PoW, PoS validators are the same local participants (prosumers, consumers or DSO) in the green energy community, which can be interconnected by a local-area network (LAN). This has two implications. On the one hand, validators can purchase green energy with the earned NRGcoins. This is a strength of PoS contributing positively to G2. On the other hand, in addition to earn transaction fees, another reason for members of the local community to perform the role of validators is because they feel responsible to maintain the operation of the NRGcoin. This is a strength of PoS with respect to G4.

5 Conclusions

In this paper, we have shown how enterprise modeling can support the strategic analysis of digitalization initiatives by using a combination of two modeling languages: SAML, for expressing strategic analyses, and ITML, for expressing IT infrastructure. Specifically, we showed how (1) SAML allows for explicitly relating strategic elements to each other, allowing for root cause analyses, and how (2) the explicit relation of SAML to ITML can be used to contextualize elements of a strategic analysis.

In terms of limitations, firstly, we notice that the produced SAML models become complex quickly. While this complexity is not a novel phenomenon (goal models tend to have a similar issue), it should be addressed since it makes the models hard to interpret, thus potentially inhibiting the added analysis capabilities started with. Secondly, in this paper we focused on modeling languages to support mainly strategic analysis of IT infrastructures. Therefore, additional extensions are required to support the strategic formulation phase in its entirety. A method to support such a model-driven analysis is also part of our future work.

References

1. Houben, G., Lenie, K., Vanhoof, K.: A knowledge-based SWOT-analysis system as an instrument for strategic planning in small and medium sized enterprises. Decision Support Systems 26(2), 125–135 (1999)
2. Hanelt, A., Hildebrandt, B., Polier, J.: Uncovering the role of IS in business model innovation - a taxonomy-driven approach to structure the field. In: European Conference of Information Systems 2015, Completed Research Papers. Paper 71 (2015)
3. Hax, A.C., Majluf, N.S.: The Strategy Concept and Process: A Pragmatic Approach. Prentice Hall, Upper Saddle River, N.J. (1996)
4. Mintzberg, H., Ahlstrand, B., Lampel, J.: Strategy Safari: A Guided Tour Through the Jungles of Strategic Management. Free Press, New York (1998)
5. Helms, M.M., Nixon, J.C.: Exploring SWOT analysis: where are we now?; a review of academic research from the last decade. JSMA 3(3), 215–251 (2010)
6. Fahy, J., Smithee, A.: Strategic marketing and the resource based view of the firm. Academy of Marketing Science Review 1 (1999)
7. Panagiotou, G.: Bringing SWOT into focus. Business Strategy Review 14(2), 8–10 (2003)

8. Bock, A., Frank, U., Bergmann, A., Strecker, S.: Towards support for strategic decision processes using enterprise models: A critical reconstruction of strategy analysis tools. In: POEM'16, pp. 41–56, Springer (2016)
9. Horkoff, J., Barone, D., Jiang, L., Yu, E., Amyot, D., Borgida, A., Mylopoulos, J.: Strategic business modeling: representation and reasoning. Software & Systems Modeling 13(3), 1015–1041 (2014)
10. Bergmann, A., Strecker, S.: Toward modeling strategic plans: Requirements & language design considerations. In et al., H.A.P., ed.: CBI 2018, Vol. 2 - Research- in-Progress Papers and Workshops, pp. 11—20, IEEE Computer Society (2018)
11. Chandler, A.D.: Strategy and Structure: Chapters in the History of the Industrial Enterprise. M.I.T. Press, Cambridge (1962)
12. IEA: Digitalization & energy. Technical report, International Energy Agency (IEA) (2017)
13. Mihaylov, M., et al.: Scanergy: a scalable and modular system for energy trading between prosumers. In: AAMAS '15, pp.1917–1918 (2015)
14. Kaczmarek-Heß, M., de Kinderen, S., Ma, Q., Razo-Zapata, I.: Modeling in support of multi-perspective valuation of smart grid initiatives. In: IEEE 12th Int. Conf. on Research Challenges in Inf. Science, 29-31 May 2018, Nantes, France, pp. 1—12, IEEE (2018)
15. Frank, U.: Outline of a method for designing domain-specific modelling languages. ICB Research Report 42, University of Duisburg-Essen, Essen (2010)
16. Pijpers, V., de Leenheer, P., Gordijn, J., Akkermans, H.: Using conceptual models to explore business-ICT alignment in networked value constellations. RE 17(3), 203–226 (2012)
17. Wegner, M.S., Hall, S., Hardy, J., Workman, M.: Valuing energy futures; a comparative analysis of value pools across UK energy system scenarios. Applied Energy 206, 815 – 828 (2017)
18. Porter, M.E.: Competitive strategy: Techniques for analyzing industries and competitors. Simon and Schuster (1998)
19. Kaplan, R.S., Norton, D.P.: Using the balanced scorecard as a strategic management system (1996)
20. Hill, T., Westbrook, R.: SWOT analysis: it's time for a product recall. Long range planning 30(1), 46–52 (1997)
21. Kurttila, M., Pesonen, M., Kangas, J., Kajanus, M.: Utilizing the analytic hierarchy process (AHP) in SWOT analysis - a hybrid method and its application to a forest- certification case. Forest Policy and Economics 1(1), 41–52 (2000)
22. Yueksel, I., Dagdeviren, M.: Using the analytic network process (AHP) in a SWOT analysis – a case study for a textile firm. Inf. Sciences 177(16), 3364–3382 (2007)
23. Phadermrod, B., Crowder, R.M., Wills, G.B.: Importance-performance analysis based SWOT analysis. International Journal of Information Management (2016)
24. Yu, E.: Towards modelling and reasoning support for early-phase requirements engineering. In: RE'97, pp. 226–235 (1997)
25. Amyot, D., Ghanavati, S., Horkoff, J., Mussbacher, G., Peyton, L., Yu, E.: Evaluating goal models within the goal-oriented requirement language. Int. J. Intell. Syst. 25(8), 841–877 (2010)
26. Overbeek, S., Frank, U., Köhling, C.: A language for multi-perspective goal modelling: Challenges, requirements and solutions. CSI 38, 1–16 (2015)
27. Horkoff, J., Aydemir, F.B., Cardoso, E., Li, T., Maté, A., Paja, E., Salnitri, M., Piras, L., Mylopoulos, J., Giorgini, P.: Goal-oriented requirements engineering: an extended systematic mapping study. Requirements Engineering (2017)
28. Gordijn, J., Yu, E., van der Raadt, B.: E-service design using i* and e3value modeling. IEEE software 23(3), 26–33 (2006)

29. Roelens, B., Steenacker, W., Poels, G.: Realizing strategic fit within the business architecture: the design of a process-goal alignment modeling and analysis technique. Software & Systems Modeling (2017)
30. Hartmann, B.: Enterprise architecture as an instrument of strategic control. In Nüttgens, M., Thomas, O., Weber, B., eds.: EMISA 2011, pp. 9-22, GI (2011)
31. Lankhorst, M.: Enterprise Architecture at Work: Modeling, Communication and Analysis. 4 edn. Springer (2017)
32. Aldea, A., Iacob, M.E., van Hillegersberg, J., Quartel, D., Bodenstaff, L., Franken, H.: Modelling strategy with AchiMate. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp.1211–1218, ACM (2015)
33. Azevedo, C.L.B., van Sinderen, M., Pires, L.F., Almeida, J.P.A.: Aligning enterprise architecture with strategic planning. In Persson, A., Stirna, J., eds.: Advanced Inf. Systems Engineering Workshops, CAiSE 2015, pp. 426–437, Springer (2015)
34. Quartel, D., Steen, M.W., Lankhorst, M.M.: Application and project portfolio valuation using enterprise architecture and business requirements modelling. Enterprise Information Systems 6(2), 189–213 (2012)
35. SoftwareAG: Aris method manual v.10 (2017)
36. Njoh, A.J.: The SWOT model's utility in evaluating energy technology: Illustrative application of a modified version to assess the sawdust cookstove's sustainability in Sub-Saharan Africa. Renew Sust Energ Rev 69, 313 – 323 (2017)
37. Razo-Zapata, I.S., Shrestha, A., Proper, E.: An assessment framework to determine the strategic value of IT architectures in smart grids. In: ACIS'17. (2017)
38. SGAM: Smart grid reference architecture. Technical report, CEN-CENELEC- ETSI Smart Grid Coordination Group, Tech. Report (2012)
39. Schuurman, P., Berghout, E.W., Powell, P.: Calculating the importance of information systems: The method of Bedell revisited. (2008)
40. Chermack, T.J., Kasshanna, B.K.: The use and misuse of SWOT analysis and implications for HRD professionals. Hum Resource Dev Int 10(4), 383–399 (2007)
41. Taghavifard, M.T., Amoozad Mahdiraji, H., Alibakhshi, A.M., Zavadskas, E.K., Bausys, R.: An extension of fuzzy SWOT analysis: An application to information technology. Information 9(3) (2018)
42. Coman, A., Ronen, B.: Focused SWOT: diagnosing critical strengths and weaknesses. International Journal of Production Research 47(20), 5677–5689 (2009)
43. Frank, U.: Multi-perspective enterprise modeling: Foundational concepts, prospects and future research challenges. SoSyM 13(3), 941–962 (2014)
44. de Kinderen, S., Kaczmarek-Heß, M.: Enterprise modeling in support of SOA migration analysis. EMISA 13(1), 1–36 (2018)
45. Bock, A., Frank, U.: Multi-perspective enterprise modeling – conceptual foundation and implementation with ADOxx. In: Domain-Specific Conceptual Modeling, Concepts, Methods and Tools. pp. 241–267, Springer (2016)
46. Xu, X., Weber, I., Staples, M., Zhu, L., Bosch, J., Bass, L., Pautasso, C., Rimba, P.: A taxonomy of blockchain-based systems for architecture design. In: 2017 IEEE International Conference on Software Architecture (ICSA). pp. 243–252, IEEE, (2017)
47. BitFury Group: Proof of stake versus proof of work. Technical report (2015)
48. ILNAS, ANEC: Blockchain and distributed ledgers: Technology, economic impact and technical standardization version 1.0. (2018)

Zwischenbetriebliche Integration in der Möbelbranche: Konfigurationen und Einflussfaktoren

Norbert Frick[1]

[1] University of Koblenz-Landau, Institute for IS Research, Koblenz, Germany
{norbert.frick}@uni-koblenz.de

Abstract. In der Möbelbranche in Deutschland existieren seit mehreren Jahren Bemühungen, zwischenbetriebliche Kooperationen zu unterstützen. In einigen Branchensegmenten (vor allem Küche und Polster) konnten sich Infrastrukturen zur Integration etablieren, die von vielen Branchenteilnehmern akzeptiert und genutzt werden. Trotz dieser Entwicklung sind überraschende Unterschiede zu beobachten: Mängel in der Prozessintegration, Unterschiede im Standardisierungsfortschritt zwischen den Branchensegmenten und die Entstehung unterschiedlicher Infrastrukturen zur zwischenbetrieblichen Integration. Die Konfigurationsanalyse nach Lyytinen und Damsgaard betrachtet Unternehmensnetzwerke als Konfigurationen zwischenbetrieblicher Integration, die einem stabilen Muster von Kooperationsformen entsprechen. Sie ist insbesondere dafür geeignet, Unternehmensnetzwerke auf Branchenebene zu untersuchen. In 21 Interviews mit 19 Organisationen wurden Daten zu Konfigurationen zwischenbetrieblicher Integration in zwei Branchensegmenten der Möbelbranche (Küche und Polster) in Deutschland erhoben und analysiert. Insgesamt wurden vier Konfigurationstypen (Mittler, Branche, Dyade und Triade) identifiziert, die in 17 konkreten Konfigurationen zum Einsatz kommen (Mittler (4-mal), Branche (5-mal), Dyade (6-mal) und Triade (2-mal)) und von 39 Einflussfaktoren geprägt sind.

Keywords: Konfigurationen, Einflussfaktoren, Interorganisationssysteme, Möbelbranche

1 Einleitung

Die Arbeitsteilung auf nationaler und internationaler Ebene nimmt in der vernetzten Welt beständig zu. Die Konzentration auf bestimmte Teile der Wertschöpfungskette führt zu einem verstärkten Einsatz von Interorganisationssystemen [1]. Zugrunde liegende Geschäftsmodelle ändern sich ebenso wie die sie unterstützenden Organisationsformen [26]. Die Dynamik der wirtschaftlichen Entwicklungen reflektiert sich insbesondere in der weiter zunehmenden Bedeutung von interorganisatorischen Informationssystemen (IOIS), die zur elektronischen Unterstützung der arbeitsteiligen Prozesse unabdingbar geworden sind.

Mit zunehmender Arbeitsteilung wächst die Heterogenität von organisatorischen und technologischen Eigenschaften, die miteinander in Einklang zu bringen sind.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Potenzielle Kooperationspartner stehen vor der Herausforderung, die für sie passende Form einer zwischenbetrieblichen Integration zu identifizieren. Mehrere Autoren suchen die Heterogenität über modellhafte Beschreibungen nachzuvollziehen (u.a. [11], [21], [7]), betrachten dabei aber meist nur Teilausschnitte der Realität (z.B. die reine Einführung, die technische Umsetzung oder die Auswirkungen eines IOIS). Die Berücksichtigung der Komplexität zwischenbetrieblicher Kooperation und der damit verbundenen zwischenbetrieblichen Integration gelingt nur selten in einem zufriedenstellenden Maß [8].

In vielen Branchen sind Phänomene zu beobachten, für die nur mit einem vertieften Verständnis der Komplexität zwischenbetrieblicher Kooperationen Erklärungsansätze zu ermitteln sind. Eine dieser Branchen ist die Möbelbranche in Deutschland. Trotz der Nutzung etablierter IOIS zur Planung und Auslieferung von hoch konfigurierbaren Produkten zeigen sich konstant hohe Reklamationsquoten von bis zu 40%[1]. Selbst auf Branchensegmentebene sind Unterschiede zu beobachten. Im Branchensegment Polster und Branchensegment Küche werden hoch konfigurierbare Produkte produziert und verkauft, trotzdem unterscheiden sie sich u.a. in der Verbandsstruktur sowie der Infrastruktur zum zwischenbetrieblichen Datenaustausch.

Es bleibt unklar, welche Formen der zwischenbetrieblichen Integration in der Möbelbranche etabliert sind und welche Eigenschaften diese Formen maßgeblich beeinflussen. Entsprechend wurden folgende Forschungsfragen formuliert: (a) *Welche Formen der zwischenbetrieblichen Integration existieren in der Möbelbranche?* und (b) *Welche Faktoren besitzen Einfluss auf die identifizierten Formen der zwischenbetrieblichen Integration?* Da es bis zum Zeitpunkt der Untersuchung keine branchenweiten Arbeiten in der Möbelbranche zu IOIS gab, bot sich die Branche als Untersuchungsraum an, um eine erste Aufnahme und Analyse von Konfigurationen zwischenbetrieblicher Integration durchzuführen.

Im weiteren Verlauf dieses Beitrags wird zunächst der Stand der Forschung zu Formen der zwischenbetrieblichen Integration und der darauf einwirkenden Einflussfaktoren beleuchtet (Kapitel 2). Im Anschluss wird das mehrstufige Forschungsvorgehen erläutert (Kapitel 3). Kapitel 4 diskutiert die Ergebnisse, während Kapitel 6 eine Zusammenfassung und kritische Würdigung enthält.

2 Formen zwischenbetrieblicher Integration

Es finden sich zahlreiche Formen zwischenbetrieblicher Integration in der Literatur, die unterschiedliche Aspekte fokussieren. Viele Gestaltdefinitionen beziehen sich auf eine oder mehrere Ebenen, in denen eine Klassifikation der von den jeweiligen Forschern wahrgenommenen Formen zwischenbetrieblicher Integration formuliert wird. Exemplarisch seien die Arbeiten von Bakos [2], Choudhury [9], Klein [13], Benasou und Venkatraman [4] und Schubert [25] erwähnt.

---

[1] Küche: https://www.moebelfertigung.com/issues/17?topic_id=103 („Fehler in der Produktion vermeiden"), LA 10.09.2018; Polster: http://www.vhk-herford.de/presse/item/moebelhandel-will-in-die-ueberholspur-wechseln/, LA 10.09.2018

Die reine Formenbestimmung zwischenbetrieblicher Integration markiert jedoch nur den Anfang einer systematischen Auseinandersetzung mit den interorganisatorischen Strukturen und den sie unterstützenden IOIS. Als typische Beschreibungsmodelle beinhalten Klassifikationsmodelle in der zwischenbetrieblichen Integration eine möglichst umfassende Kollektion aller möglichen Ausprägungen eines IOIS. Bekannte Ansätze wurden unter anderem von Barrett und Konsynski [3], Klein [13] oder Hong [12] entwickelt. Sie konzentrieren sich vor allem auf die Klassifikation eines IOIS unter Berücksichtigung verschiedener Kriterien, die sich üblicherweise auf technologische, organisatorische und/ oder institutionelle Eigenschaften beziehen. Allerdings fehlt diesen Modellansätzen in der Regel jegliche Entwicklungsperspektive, da sich aus den Charakteristika eines IOIS nicht zwangsläufig die zu deren Entstehung führenden Einflussfaktoren und Veränderungsmechanismen ableiten lassen.

2.1 Einflussfaktoren auf Formen zwischenbetrieblicher Integration

Typische Formen zwischenbetrieblicher Integration, die in den zuvor genannten Arbeiten aber auch in vielen anderen Beiträgen thematisiert werden, sind von Faktoren auf institutioneller, organisatorischer und technischer Ebene abhängig. Wenngleich manche Arbeiten solche Faktoren nicht explizit erwähnen, so lassen sie sich aus dem jeweiligen Kontext heraus ableiten. Rodon und Sese [22] charakterisieren z.B. vier IOIS-Typen mit Hilfe der beiden Einflussfaktoren Macht und Regeln. Die Autoren beziehen sich hauptsächlich auf die beiden vorgenannten Differenzierungsmerkmale, indem sie anhand der Machtverhältnisse und Regelwerke die vier Typen beschreiben. Trotzdem fließt z.B. im Typ „Market" das Charakteristikum „Wettbewerb" mit ein, während im Typ „Hub" die zwischenbetrieblichen Beziehungen und die darauf basierenden Vertragsgestaltungen zusätzlich eine Rolle spielen.

Implizite Untersuchungen bzw. Analysen von Einflussfaktoren sind in vielen Beiträgen zu finden. Eine gemeinsame konzeptionelle Betrachtung von Formen zwischenbetrieblicher Integration und ihrer jeweiligen Einflussfaktoren gibt es hingegen nur selten. Meist findet man aus Einzelperspektiven heraus motivierte Untersuchungen (z.B. Information Sharing [3], IOIS Implementation [6], oder Transaction Costs [10]), die punktuell einzelne Einflussfaktoren herausgreifen und im Rahmen ihrer Perspektive untersuchen. Lyytinen und Damsgaard [16] schlagen in ihrem Positionspapier die Konfigurationsperspektive als adäquate analytische Linse vor, um einen ersten umfassenderen analytischen Blick auf Integrationsformen und ihre Einflussfaktoren zu gewinnen.

2.2 Konfigurationen zwischenbetrieblicher Integration

Lyytinen und Damsgaard [16] verstehen unter IOIS-Konfigurationen organisierte Cluster von IOIS-Nutzern, die durch ausgewählte Einflussfaktoren, sogenannte Schlüssel-Elemente, beeinflusst werden. Dabei berufen sie sich auf ähnliche Einsätze des Konfigurationsbegriffs in der Physik, in der Organisationsforschung oder in Service Science, die aus ihrer Sicht auf die IOIS-Domäne übertragbar sind. Eine bekannte Definition entstammt dem Buch von Miller und Friesen [18], die Konfigurationen auf den Organisationskontext beziehen: „These can be defined as commonly occurring clusters of attributes or relationships […] that are internally cohesive, such that the presence of some attributes suggest the reliable occurrence of others.” Die Grundlage einer Konfiguration bilden dementsprechend nicht nur die einzelnen Faktoren (attributes or relationships), sondern auch die zwischen ihnen existierenden Abhängigkeiten (internally cohesive). Meyer et al. [17] betonen dabei „coherent patterns“, schlüssige Muster, die sich aus den Interdependenzen der einzelnen Attribute ergeben. Konfigurationen basieren also nicht auf einem zufälligen Zusammenkommen einzelner Faktoren, sondern reflektieren eine begrenzte Anzahl stabiler Zustände [16].

Die von den Autoren beschriebenen Konfigurationstypen (Dyade, Hub and Spoke, Industry und Community) und die sie beschreibenden Schlüssel-Elemente (organizing vision, key functionality, mode of interaction, structure und mode of appropriation) stellen eine Auswahl der von ihnen im Laufe ihrer Forschungsarbeiten vorgefundenen Konfigurationen und Einflussfaktoren dar. Diese Auswahl ist vor dem Hintergrund bereits existierender Arbeiten nicht vollständig, bietet aber als eine der ersten Arbeiten einen konzeptionellen Rahmen für eine ganzheitliche Untersuchung zwischenbetrieblicher Integrationsformen und der auf sie einwirkenden Einflussfaktoren. Daher wurde die von Lyytinen und Damsgaard vorgeschlagene Konfigurationsperspektive als analytische Linse zur Beantwortung der in Kapitel 1 gestellten Forschungsfragen angewendet.

3 Forschungsansatz

Die Diskussion der Konfigurationsperspektive im vorangegangenen Kapitel zeichnet ein multiperspektivisches Bild der gegenwärtigen Betrachtungsweisen von Unternehmensnetzwerken und die sie unterstützenden IOIS. Es reicht nicht mehr aus, sich auf ein IOIS selbst als Untersuchungsgegenstand zu fokussieren. Die Forschungsfragen (siehe Kapitel 1) zielen auf ein möglichst ganzheitliches Verständnis der technischen, organisatorischen und institutionellen Eigenschaften von IOIS-unterstützten Unternehmensnetzwerken ab.

In der Konsequenz sind als Untersuchungsgegenstand die Konfigurationen zwischenbetrieblicher Integration selbst zu wählen. Die Konfigurationsperspektive erlaubt die Verwendung einer in sich geschlossenen Logik und Sprache zur Beschreibung/ Modellierung und ganzheitlichen Untersuchung von IOIS-gestützten Unternehmensnetzwerken. Als Untersuchungsbereich werden gemäß der in Kapitel 1 beschriebenen Motivation die beiden Branchensegmente Küche und Polster gewählt. In beiden Branchensegmenten gibt es gegenwärtig kaum öffentlich zugängliche Informationen über Form und Eigenschaften der darin aktiven IOIS-gestützten Unternehmensnetzwerke.

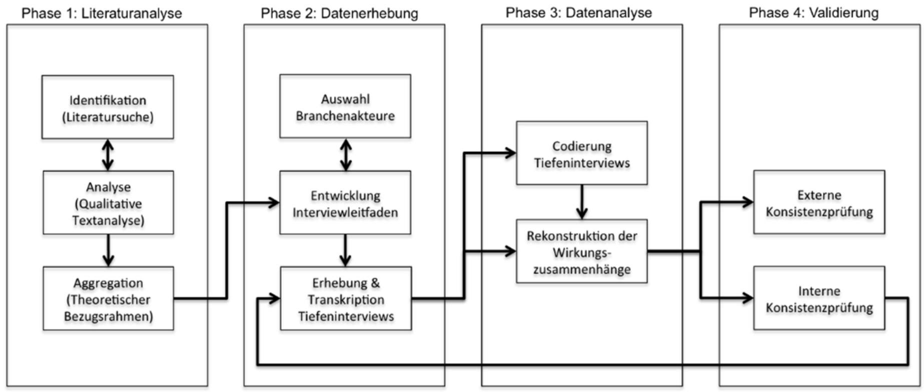


Abbildung 1: Forschungsvorgehen zur Konfigurationsanalyse

Die Konfigurationsanalyse nach Lyytinen und Damsgaard [16] bietet keine vordefinierten Verfahrensweisen für ein Forschungsvorhaben, sondern nur eine auf Basis eines gewählten Forschungsinteresses begrenzte Anzahl an Vorschlägen. Der explorative Charakter des Forschungsprojekts legt nahe, dass die genutzten Forschungsmethoden flexibel genug sein müssen, um dem Forscher genügend Freiraum zur Kreativität in der Informationsbeschaffung zu lassen [5]. Gemäß der Problemstellung (vgl. Kapitel 1) fehlt es an methodischen und inhaltlichen Informationen, um die gestellten Forschungsfragen zur Möbelbranche sinnhaft beantworten zu können. Folglich ist eine Datenerhebung in der einschlägigen wissenschaftlichen Literatur zur Abgrenzung des Konfigurationsbegriffs sowie der Operationalisierung der Konfigurationsanalyse ebenso notwendig, wie die empirisch belegte Bestimmung von Konfigurationen in den beiden zu untersuchenden Branchensegmenten. Dementsprechend wurde ein mehrphasiges Forschungsvorgehen auf Basis der Konfigurationsanalyse entwickelt, das vier Hauptphasen umfasst (vgl. Abbildung 1):

- Phase 1: Literaturanalyse ((1) Identifikation und Beschaffung von relevanten Literaturbeiträgen zu Konfigurationen zwischenbetrieblicher Integration, (2) Analyse der gefundenen Literaturbeiträge hinsichtlich Konfigurationstypen und Einflussfaktoren, (3) Ableitung eines Bezugsrahmens zur Weiterverwendung in der Phase der Datenerhebung, (4) Deskriptive Codierung der vorgefundenen Beiträge)
- Phase 2: Datenerhebung ((1) Entwicklung eines semi-strukturierten Interviewleitfadens, (2) Auswahl von relevanten Akteuren mittels „Snowball or chain sampling" [20], (3) Durchführung von 21 Tiefeninterviews mit 19 Unternehmen/ Organisationen [19].
- Phase 3: Datenanalyse ((1) Deskriptive und Strukturierte Codierung der Interviewtransskripte [23], (2) Durchführung von der Codierungszyklen, (3) Auswertung von mehr als 550 identifizierten Codes)
- Phase 4: Validierung ((1) Validierung von Aussagen mittels inhaltlicher Triangulation ([19], [20]), (2) Validierung mit begleitenden Dokumenten)

4 Konfigurationen zwischenbetrieblicher Integration in den Branchensegmenten Küche und Polster

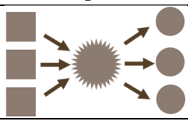
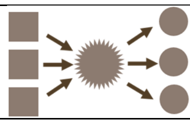
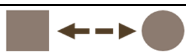
Insgesamt können 39 Einflussfaktoren aus der Analyse der Codierung der 21 Interviews identifiziert werden, wobei 7 Faktoren (Marktdominanz der Filialisten, Wettbewerb Eigenmarken (Händler), Wettbewerb Eigenmarken (Verbände), Barriere zur Vertikalisierung der Hersteller, Verbandsdienstleistungen Verbände, Stammdatenverwaltung Verbände und Organisation Intermediäre) in beiden Branchensegmenten eine Rolle spielen. Im Küchensegment können 18 Einflussfaktoren den Konfigurationen zugeordnet werden, während es im Branchensegment Polster 28 Faktoren sind. Die in beiden Branchensegmenten vorgefundenen Einflussfaktoren und Konfigurationen zwischenbetrieblicher Integration lassen sich zusätzlich auf unterschiedlichen Analyseebenen identifizieren und diskutieren: Ebene der *Handelsbeziehungen*, Ebene der *Technischen Realisierung* und Ebene der *Standardisierung*.

4.1 Ebene der Handelsbeziehungen

Es lassen sich zwei Typen von Daten identifizieren, die zwischen den Akteuren auf Ebene der Handelsbeziehungen in beiden Branchensegmenten ausgetauscht werden müssen: Stammdaten zum Austausch von Informationen über das zwischen den Handelspartnern vereinbarte Sortiment (vor allem Katalogdaten), sowie Bewegungsdaten zum Austausch von Geschäftsbelegen, wie z.B. einem Auftrags- oder Rechnungsdokument. Ein Händler verhandelt mit einem Hersteller das von diesem zu liefernde Sortiment inklusive der dazugehörigen Konditionen und bezieht nach Vertragsschluss die entsprechenden Katalogdaten vom Hersteller. Auf deren Basis wird eine Küche bzw. ein Polstermöbel für den Endkunden beim Hersteller bestellt, dort produziert, zum Händler geschickt und vom Händler beim Endkunden aufgebaut. Die Händler in beiden Branchensegmenten sind in sogenannten Einkaufsverbänden organisiert, die v.a. Preisvorteile gegenüber den Herstellern realisieren sollen. Über die Jahrzehnte hinweg haben sich die Einkaufsverbände zu Organisationsverbänden weiterentwickelt, die nicht nur günstige Einkaufskonditionen erzielen, sondern ihren Mitgliedern weitere Dienstleistungen, wie z.B. ein Produktkatalogmanagement, anbieten. Sie dienen in vielen Fällen als Mittler zwischen Herstellern und ihren angeschlossenen Händlern.

Tabelle 1: Topologie der Konfigurationen auf Ebene der Handelsbeziehungen

		Branchensegment Küche		Branchensegment Polster	
H	S		**D**		**D**
		Einflussfaktoren Küche: W: Marktdominanz Küchenmöbelindustrie			

		Einflussfaktoren Polster: W: Eigenmarken (Händler), W: Eigenmarken (Verbände), W: Wettbewerb PKS, S: Fachkräftemangel (Filialisten), K: Multi-Channel (Kunden), K: Informationstransparenz (Kunden), K: Beratung (Kunden), K: Informationstransparenz (Kunden)			**D**
		Einflussfaktoren für beide Branchensegmente: W: Marktdominanz Filialisten; W: Barriere Vertikalisierung Hersteller, F: Produktplanung; P: Produktkomplexität; P: Preiskomplexität			
			**M**		**M**
		Einflussfaktoren Küche: *keine speziellen*			**M**
		Einflussfaktoren Polster: W: Eigenmarken (Händler), W: Eigenmarken (Verbände)			
		Einflussfaktoren für beide Branchensegmente: W: Barriere Vertikalisierung Hersteller, W: Marktdominanz Filialisten, W: Stammdatenverwaltung Verbände, W: Eigenmarken (Verbände)			
	Bewegungsdaten		**D**		**D**
		Einflussfaktoren Küche: W: Marktdominanz Küchenmöbelindustrie			**D**
		Einflussfaktoren Polster: W: Eigenmarken (Händler) W: Wettbewerb PKS, S: Fachkräftemangel (Filialisten), K: Multi-Channel (Kunden), K: Informationstransparenz (Kunden), K: Beratung (Kunden), K: Informationstransparenz (Kunden)			
		Einflussfaktoren für beide Branchensegmente: W: Marktdominanz Filialisten; W: Barriere Vertikalisierung Hersteller, F: Produktplanung; P: Produktkomplexität; P: Preiskomplexität			

(1) Kategorien von Einflussfaktoren: W=Wettbewerb, M=Macht, P=Produkteigenschaften, B=Beziehungen, S=Struktur, F=Funktionen, K=Kundenanforderungen
(2) Konfigurationstypen: D=Dyade, M=Mittler

Auf Ebene der Handelsbeziehungen finden sich in beiden Branchensegmenten die gleichen Konfigurationstypen: Die *Dyade* als Repräsentant von bilateralen Austauschbeziehungen von Stamm- und Bewegungsdaten zwischen Herstellern und Händlern (D) und die *Mittler*-Konfiguration als Repräsentant von Hub-and-Spoke ähnlichen Strukturen (M), in denen ein zentraler Akteur die Koordination des Datenaustauschs übernimmt (vgl. Tabelle 1). Der Grund dafür liegt in den ähnlichen Geschäftsmodellen und Akteurskonstellationen in beiden Branchensegmenten, die sich in den genannten stabilen Austauschbeziehungen manifestieren.

Betrachtet man die Einflussfaktoren auf die jeweiligen Konfigurationen, so zeigten sich ebenfalls kaum Unterschiede. Der Wettbewerb zwischen den Branchenakteuren im Küchensegment ist weitestgehend stabil, während für dyadische Konfigurationen die strukturellen Voraussetzungen vor allem bei großen Unternehmen gegeben sind, die deren Weiterentwicklung im Polstersegment aus Eigeninteressen massiv vorantreiben. Dieser Umstand reflektiert sich auf Ebene der Handelsbeziehungen im Einflussfaktor „Wettbewerb zwischen Herstellern von Polsterkonfigurationssystemen (PKS)“, da die elektronisch gestützte Produktplanung im Gegensatz zum Küchensegment noch nicht dauerhaft im Polstersegment etabliert ist. Zusätzliche Faktoren aus dem Bereich Kundenanforderungen spielen ebenfalls eine Rolle, wenngleich deren Bedeutung für die Entstehung und Stabilisierung von zwischenbetrieblichen Kooperationsmustern untergeordnet ist.

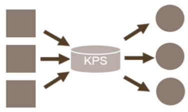
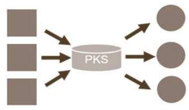
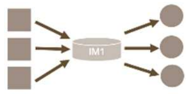
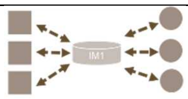
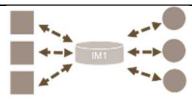
4.2 Ebene der Technischen Realisierung

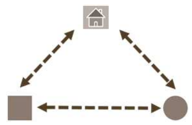
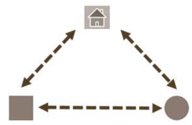
Branchensegment Küche. Auf der Ebene der Technischen Realisierung kommen im Küchensegment drei weitere Akteure hinzu, die im zwischenbetrieblichen Datenaustausch eine Rolle spielen: (1) Die KPS (Küchenplanungssoftware)-Anbieter nehmen im Branchensegment Küche eine zentrale Rolle in der technischen Realisierung ein. Sie unterstützen bei der elektronischen Planung von Küchen, die aufgrund der vielen unterschiedlichen zu beachtenden Rahmenparameter hochkomplex ist. Die KPS-Anbieter lassen sich direkt beim Handel verorten, da die von ihnen entwickelten Planungslösungen in erster Linie an Handelshäuser vertrieben werden und die KPS die von den Herstellern zugelieferten Katalogdaten verarbeiten. In der Beziehung zwischen Herstellern, Verbänden und Händlern zeigt sich ein ähnliches Bild. So werden KPS bei jedem einzelnen Händler direkt betrieben und verwalten die bei ihm eingespielten Herstellerkataloge zur Küchenplanung. Ein Verband hat i.d.R. keine KPS-Lösung im Einsatz. (2) Der Intermediär IM1 bietet Dienstleistungen zum Katalog- und Bewegungsdatenaustausch zwischen Herstellern und Händlern in beiden Branchensegmenten an, wobei im Branchensegment Küche nur Bewegungsdaten und im Branchensegment Polster sowohl Katalog- als auch Bewegungsdaten über die intermediale Plattform ausgetauscht werden. Das Angebot wird sowohl von Filialisten als auch von kleineren Händlern genutzt. (3) IT-Dienstleister mit dem Schwerpunkt Zentralregulierung bieten gerade kleineren Händlern einen signifikanten Effizienzgewinn dank der organisatorischen Abwicklung der Zahlungen zwischen ihnen und den angebundenen Herstellern als unverzichtbare Dienstleistung. Sie nehmen keinen besonderen Einfluss auf andere Konfigurationen.

Branchensegment Polster. Im Gegensatz zum Küchensegment spielen auf Ebene der Technischen Realisierung andere Akteure eine Rolle. (1) Softwareanbieter von Polsterkonfigurationssystemen (PKS) versuchen bereits seit etwa 10 Jahren, grafische Konfigurationslösungen im Handel zu platzieren. Es ist aber erst in den letzten zwei bis drei Jahren gelungen, ernsthafte Fortschritte in der Datenverfügbarkeit für die Lösungen zu erzielen. Dafür werden vor allem Anforderungen aus dem Handel verantwortlich gemacht (z.B. Informationstransparenz für Endkunden, steigender Fachkräftemandel), die eine entsprechende Nachfrage nach PKS beflügeln. (2) Im Branchensegment Polster sind gegenwärtig zwei Intermediäre aktiv, die ihre Dienstleistungen anbieten: IM1 und IM2. IM1 bietet den Austausch von Bewegungsdaten zwischen Herstellern und Händlern an, verteilt über seine Plattform im Gegensatz zum Branchensegment Küche aber auch Katalogdaten von einigen Herstellern an den Handel. Damit besitzt IM1 als neutraler Datenmittler im Branchensegment Polster eine größere Bedeutung als im Branchensegment Küche. IM2 hingegen bietet keine Plattform zum Datenaustausch an, sondern realisiert zwischen je zwei Geschäftspartnern (Hersteller und Händler) eine bidirektionale Integration, die tief in die jeweiligen Produktionssysteme (Hersteller) bzw. Warenwirtschaftssysteme (Händler) integriert ist. Dadurch ist eine nahezu vollständige

Daten- und Anwendungsintegration möglich, über die sowohl Katalog- als auch Bewegungsdaten ausgetauscht werden können. Die Integration wird über ein proprietäres Datenformat realisiert, das den Mitbewerbern nicht zur Verfügung steht.

Tabelle 2: Topologie der Konfigurationen auf Ebene der Technischen Realisierung

		Branchensegment Küche		Branchensegment Polster	
Technische Realisierung	Stammdaten		**M**		**M**
		Einflussfaktoren Küche: W: Geschäftsmodell KPS, W: Wettbewerb KPS			**M**
		Einflussfaktoren Polster: W: Geschäftsmodell PKS, W: Wettbewerb PKSS: Fachkräftemangel (Filialisten), M: Einflussnahme Filialist auf IM2, M: Unabhängigkeit von SW-Handel, W: Datenvollständigkeit (Filialist), B: Initiative Standardisierung, M: Partnerabhängigkeit (Hersteller), W: Marktdominanz Filialisten			
		Einflussfaktoren für beide Branchensegmente: F: Produktplanung; P: Produktkomplexität; P: Preiskomplexität			
		-			**B**
		Einflussfaktoren Küche: *keine*			**B**
		Einflussfaktoren Polster: F: Funktion Polsterkonfiguration, P: Produktkomplexität Polster, P: Preiskomplexität Polster, W: Wettbewerb PKS, S: Fachkräftemangel (Filialisten), W: Geschäftsmodell IM1			
		Einflussfaktoren für beide Branchensegmente: *keine*			
		-			**D**
		Einflussfaktoren Küche: *keine*			**D**
		Einflussfaktoren Polster: W: Geschäftsmodell IM2, W: Kritische Masse IM2, B: Strat. Allianz SW-Handel & IM2, B: Strat. Allianz IM2 & Verband, B: Initiative Datenintegration, M: Partnerabhängigkeit (Hersteller), W: Marktdominanz Filialisten, F: Funktion Küchenplanung, P: Produktkomplexität Polster, P: Preiskomplexität Polster, W: Wettbewerb PKS, S: Fachkräftemangel (Filialisten)			
		Einflussfaktoren für beide Branchensegmente: *keine*			
	Bewegungsdat				**B**
		Einflussfaktoren Küche: B: Strategische Allianz IM1/ nmedia			**B**
		Einflussfaktoren Polster: W: Geschäftsmodell IM1			
		Einflussfaktoren für beide Branchensegmente: S: Organisation IM1			

					**T**
		Einflussfaktoren Küche: *keine*			**T**
		Einflussfaktoren Polster: *keine*			
		Einflussfaktoren für beide Branchensegmente: Verbandsdienstleistungen Verbände			
		-			**D**
		Einflussfaktoren Küche: *keine*			**D**
		Einflussfaktoren Polster: W: Geschäftsmodell IM2, W: Kritische Masse IM2, B: Strat. Allianz SW-Handel & IM2, B: Strat. Allianz IM2 & Verband, B: Initiative Datenintegration, M: Partnerabhängigkeit (Hersteller), W: Marktdominanz Filialisten			
		Einflussfaktoren für beide Branchensegmente: *keine*			
(1) Kategorien von Einflussfaktoren: W=Wettbewerb, M=Macht, P=Produkteigenschaften, B=Beziehungen, S=Struktur, F=Funktionen, K=Kundenanforderungen (2) Konfigurationstypen: B=Branche, D=Dyade, M=Mittler, T=Triade					

Im Rahmen des Stammdatenaustausches liefern die für die Realisierung der zwischenbetrieblichen Integration zuständigen KPS-Dienstleister nicht nur Lösungen zur Umsetzung der Konfigurationen auf Ebene der Handelsbeziehungen, sondern nehmen aufgrund unterschiedlicher Faktor-Ausprägungen im Branchensegment Küche eine gestaltende Rolle ein. Diese Besonderheit reflektieren 21 Codes in der Kategorie „Wettbewerb“, die unterschiedliche Einflussfaktoren mit Bezug zu Softwareanbietern interpretieren. Besonders ist eine KPS-Konfiguration zu nennen (31 Erwähnungen in 16 Interviews), die als Mittler-Konfigurationstyp (M) die Bereitstellung von Katalogdaten für den Handel organisiert. Bei dieser Konfiguration liegt keine Hub-and-Spoke-Konfiguration im Sinne von Lyytinen und Damsgaard [16] vor. Die KPS-Anbieter werten die Daten massiv auf, indem sie dreidimensionale Produktmodelle und fotorealistisches Aussehen der Küchenplanung hinzufügen. Die Weiterleitung an die Händler ist hier nicht von einer Verbandszugehörigkeit abhängig, sondern davon, ob der jeweilige Händler Kunde des KPS-Herstellers ist. Insofern ist „KPS-Sortiment“ ein geschlossenes System.

Im Branchensegment Polster (Fokus Stammdaten) sind deutlich mehr Konfigurationen zwischenbetrieblicher Integration zu beobachten als im Branchensegment Küche. Die erste Konfiguration entspricht der KPS-Konfiguration und wird insgesamt 20-mal in den Interviews beschrieben. Dieser Mittler-Konfigurationstyp (M) wird von den PKS-Anbietern dazu genutzt, Katalogdaten aus der Industrie dem Handel zur Verfügung zu stellen, wobei die Katalogdaten entsprechend mit Grafiken aufgewertet werden. Der Konfigurationstyp hat eine deutlich geringere Verbreitung, da sich die PKS-Anbieter noch nicht in gleicher Weise im Branchensegment Polster als Datenmittler etabliert haben, wie die KPS-Anbieter im Branchensegment Küche. Die Branchen-Konfiguration (B), die durch IM1 realisiert wird, spielt kaum eine Rolle, da nur vereinzelt Kataloge übertragen werden. IM1 wird hauptsächlich für die Datenübertragung von Fachsortimentsdaten genutzt.

Besonders hervorzuheben sind die beiden dyadischen Konfigurationen (D) des Intermediärs IM2 (Stamm- und Bewegungsdaten), die aufgrund des proprietären Datenformats geschlossene Konfigurationen darstellen. Die Funktion der proprietären Schnittstellentechnik von IM2 entspricht der eines Konverters. Die IM2-Konfigurationen sind bei vielen Filialisten sowie den großen Polsterherstellern in Deutschland gegenwärtig im Einsatz.

Die Konfigurationen auf Basis von Bewegungsdaten unterscheiden sich kaum voneinander, da der Bedarf z.B. für den Rechnungsaustausch in beiden Segmenten gegeben ist. Ein Beispiel dafür ist die Triade-Konfiguration (T), die ein verbandstypisches Phänomen der Delkredere-Übernahme zur Absicherung der Zahlungsunfähigkeit von Verbandsmitgliedern beschreibt. Obwohl diese Konfiguration nur innerhalb von Organisationsverbänden eine Rolle spielt, komplettiert sie die Übersicht aller Konfigurationen, da sie für viele kleinere Handelsunternehmen ein integraler Bestandteil für deren Geschäftstätigkeiten darstellt.

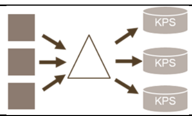
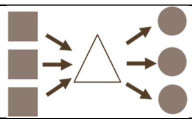
4.3 Ebene der Standardisierung

Als zusätzlicher Akteur auf Ebene der Standardisierung im Branchensegment Küche ist das SG1 zu nennen. Das SG1 fungiert als Standardisierungsgremium u.a. für die beiden Branchensegmente Küche und Polster. Seine Aufgabe besteht darin, neue Versionen der ihm zugeordneten Datenstandards zu definieren und zu verabschieden. Mit der Standardisierung der einzelnen Formate versprechen sich die Hersteller vor allem Effizienzvorteile in der Bearbeitung von Handelsbestellungen. Die Reklamationsquoten in den beiden Branchensegmenten sind im Schnitt mit etwa 40% immer noch sehr hoch. Fehler sind durchaus auf der Herstellerseite in Produktionsmängeln zu finden bzw. in der Logistik. Einen großen Anteil der Reklamationen machen jedoch auch Fehler in den Handelsbestellungen aus.

Die Initiative zur Gründung des SG1 ist von der Möbelindustrie ausgegangen, die sich als Datenlieferant für den Möbelhandel versteht. Dieses Selbstverständnis hat dazu geführt, dass das SG1 sich nicht nur in der Datenstandardisierung engagiert, sondern auch Dienstleistungen als Datenmittler zur Verfügung stellt. Dazu existiert eine Kooperation mit einem IT-Dienstleister der Industrie, der für das SG1 einen Stammdatenserver betreibt. Auf diesen Server laden gegenwärtig fast alle Hersteller, die in Deutschland über Händler Küchen vertreiben, ihre Katalogdaten hoch und stellen sie dem Handel bzw. KPS-Anbietern kostenfrei zur Verfügung. Dabei können sowohl die KPS-Anbieter als auch große Händler (Filialisten) Katalogdaten direkt vom Stammdatenserver beziehen. Handelstransaktionen werden nicht über das zentrale SG1-Angebot abgewickelt. Dazu kommen weiterhin die bereits diskutierten bilateralen Schnittstellen oder IM1 als Intermediär zum Einsatz. Der Stammdatenserver wird finanziell vom SG1 getragen. Der Verband spielt im Zusammenhang mit dem SG1 kaum eine Rolle. In dieser Konstellation werden von den einzelnen Verbandshändlern i.d.R. keine Daten vom Stammdatenserver des SG1 heruntergeladen, sondern von den KPS-Anbietern den einzelnen Mitgliedern direkt zur Verfügung gestellt. Insbesondere die kleineren Händler sind in diesem Fall auf die Datenversorgung durch die KPS-

Anbieter angewiesen, da sie i.d.R. keine eigene oder nur eine begrenzte Stammdatenpflege betreiben.

Tabelle 3: Topologie der Konfigurationen auf Ebene der Standardisierung

		Branchensegment Küche		Branchensegment Polster	
Standardisierung	Stammdaten		**B**	-	
		Einflussfaktoren Küche: W: Marktdominanz Filialisten, W: Eigenmarken (Händler), W: Eigenmarken (Verbände), W: Informationsintransparenz (Händler), W: Informationsintransparenz (Verbände), F: Funktion Reklamation (Händler), W: Wettbewerb KPS, M: Blockadehaltung (Technologie)			**B**
		Einflussfaktoren Polster: *keine*			
		Einflussfaktoren für beide Branchensegmente: *keine*			
			**B**	-	
		Einflussfaktoren Küche: W: Marktdominanz Filialisten, W: Eigenmarken (Händler), W: Eigenmarken (Verbände), W: Informationsintransparenz (Händler), W: Informationsintransparenz (Verbände), F: Funktion Reklamation (Händler)			**B**
		Einflussfaktoren Polster: *keine*			
		Einflussfaktoren für beide Branchensegmente: *keine*			

(1) Kategorien von Einflussfaktoren: W=Wettbewerb, M=Macht, P=Produkteigenschaften, B=Beziehungen, S=Struktur, F=Funktionen, K=Kundenanforderungen
(2) Konfigurationstypen: B=Branche, D=Dyade, M=Mittler, T=Triade

In der ersten Branchen-Konfiguration (B) im Branchensegment Küche versorgen sich die aktiven KPS-Anbieter regelmäßig mit den aktuellen Herstellerkatalogen im IDM-Küche Format, die sie in ihren Küchenplanungssystemen einpflegen. Diese Konfiguration entspricht einer Branchenkonfiguration nach Lyytinen und Damsgaard [16], da sie für alle registrierten Branchenteilnehmer einen offenen Zugang zu den Katalogen der Hersteller darstellt. Die Idee eines öffentlichen Guts [24] für alle Akteure in der Branche wird durch den registrierungspflichtigen, kostenfreien Zugang und die Informationstransparenz in der Standarddokumentation unterstrichen.

Die zweite vom SG1 gestützte Branchen-Konfiguration (B) sieht nicht die KPS-Anbieter, sondern die Händler in der Rolle des Datenbeziehers. Diese Konfiguration spielt jedoch gegenwärtig kaum eine Rolle, da die Verarbeitung und Bereitstellung der Daten hauptsächlich über die KPS-Anbieter stattfindet. Es gibt bislang (Stand: September 2018) nur drei gelistete Händler, die am Stammdatenserver registriert sind und Katalogdaten beziehen.

Im Branchensegment Polster gibt es im Gegensatz zum Küchensegment gegenwärtig keine Konfiguration zwischenbetrieblicher Integration auf der Ebene der Standardisierung. Das SG1 bietet die Leistungen des Stammdatenservers nur für den

IDM-Küche-Standard an. Die Verbreitung von Katalogdaten im IDM-Polster-Format wird aktuell den jeweiligen Branchenakteuren selbst überlassen.

5 Diskussion der Ergebnisse

Akteursrollen: Die vorliegende Untersuchung liefert nicht nur ein detaillierteres Bild einzelner Konfigurationen, sondern hebt die Bedeutung einzelner Akteursrollen innerhalb und zwischen den Konfigurationen hervor Dies zeigt sich sehr deutlich am Beispiel der IT-Dienstleister in der Branche, die für sieben der 12 Konfigurationen verantwortlich zeichnen. Die zwischen ihnen herrschenden wettbewerblichen Bedingungen und die darauf resultierenden Systementwicklungen sind teilweise für branchendominante Konfigurationen verantwortlich. In den identifizierten Literaturbeiträgen der Konfigurationsforschung wird IT-Dienstleistern als branchenrelevante Akteure jedoch kaum Bedeutung beigemessen. Die von IT-Dienstleistern bereitgestellten Konfigurationen werden auf die jeweiligen Handelspartner in der Regel implizit übertragen: IOIS werden als infrastrukturelle Elemente betrachtet, die losgelöst von anderen Einflussfaktoren den Datenaustausch realisieren. In der vorliegenden Arbeit zeigt sich, dass der Wettbewerb zwischen IT-Dienstleistern und die daraus entstehenden unterschiedlichen Konfigurationen sehr wohl eine signifikante Rolle spielen können. Es ist also in künftigen Untersuchungen zu klären, anhand welcher Kriterien eine Konfiguration auf der Ebene der Technischen Realisierung anzusiedeln ist und damit ein Einfluss der Akteursrolle IT-Dienstleister auf die Branche bzw. auf die Konfigurationen in einer Branche anerkannt wird.

Analyseebenen: Die herausgearbeiteten inhaltlichen Schwerpunkte (Handelsbeziehungen, Technische Realisierung und Standardisierung (Analyseebenen), sowie Stamm- und Bewegungsdaten (Datentypen)) sind auf andere Branchen entsprechend der dort herrschenden Akteurskonstellationen und der Art der Datenintegration anzupassen bzw. zu ersetzen. So kann z.B. die Ebene der Standardisierung entfallen, wenn es keine Bestrebungen zu einer branchenübergreifenden Formatentwicklung gibt. Auf der anderen Seite können u.U. nicht Stammdaten, sondern strategische Informationen (z.B. Echtzeitdaten zur Kundennachfrage) für den Datenaustausch relevant sein [15]. Der Fokus der beiden Analyseperspektiven ist somit abhängig vom Untersuchungsgegenstand und -ziel zu wählen.

Generalisierbarkeit der Ergebnisse: Im direkten Vergleich der beiden Branchensegmente auf den drei Ebenen Handelsbeziehungen, Technische Realisierung und Standardisierung zeigt sich, dass die Rolle und das Selbstverständnis der Branchenakteure maßgeblich das Zustandekommen von Konfigurationen zwischenbetrieblicher Integration beeinflussen. Die Konfigurationen stammen aus unterschiedlichen Motiven der Akteure heraus, die teilweise wirtschaftliche, teilweise aber auch normative Hintergründe besitzen. Die Einflussfaktoren für das Zustandekommen sind somit als branchenspezifisch zu bewerten. Eine Generalisierung der Konfigurationen auf andere Branchen(segmente) ist insofern möglich, als dass sich die Konfigurationen selbst als modellhafte Repräsentation der zwischenbetrieblichen

Zusammenarbeit in vielen unterschiedlichen Bereichen wiederfinden lassen. So haben Lyytinen und Damsgaard die Konfigurationstypen „Dyade“ und „Branche“ bereits in früheren Untersuchungen identifiziert. Konfigurationen können somit als gestaltende Elemente einer Modellierung von interorganisationalen Kooperationen genutzt werden.

6 Zusammenfassung, Kritische Würdigung und Ausblick

Die vorliegende Arbeit hat als eine der ersten Arbeiten die von Lyytinen und Damsgaard [16] konzeptionell vorgeschlagene Konfigurationsanalyse operationalisiert und in zwei Branchensegmenten (Küche und Polster in der Möbelbranche) die Forschungsfragen (a) *Welche Formen der zwischenbetrieblichen Integration existieren in der Möbelbranche?* und (b) *Welche Faktoren besitzen Einfluss auf die identifizierten Formen der zwischenbetrieblichen Integration?* beantwortet. Mittels eines explorativen Forschungsansatzes wurden 21 Tiefeninterviews mit 19 Organisationen geführt. Die vorgefundenen 17 Konfigurationen und die auf sie einwirkenden 39 Einflussfaktoren wurden systematisch untersucht. Es konnten vier Konfigurationstypen (Branche, Dyade, Mittler und Triade) sowie sieben Kategorien von Einflussfaktoren (Wettbewerb, Macht, Produkteigenschaften, Beziehungen, Struktur, Funktionen, Kundenanforderungen) identifiziert werden.

Die Untersuchung bietet trotz des tieferen Einblicks in die zwischenbetrieblichen Kooperations- und Integrationsbeziehungen keine vollumfängliche Branchenanalyse. Selbst mit einer sorgfältigen Auswahl der Interviewpartner (im Hinblick auf die einbezogenen Perspektiven und Rollen der Akteure) ist nur ein eingeschränkter Blick in die Branche möglich. Die Ergebnisse sind folglich in künftigen Untersuchungen in der Branche empirisch zu validieren.

Die Analyse der Konfigurationen in den Branchensegmenten Küche und Polster zeigt, dass ein in der zwischenbetrieblichen Integration engagierter Akteur in der Regel an mehreren Konfigurationen beteiligt ist, die darüber hinaus in Beziehung zueinanderstehen. Dieses Phänomen der geschachtelten Konfigurationen bietet sich als weiterer Untersuchungsgegenstand an, der in künftigen Forschungsprojekten untersucht werden sollte.

Literatur

1. Alter, S.: System interaction theory: Describing interactions between work systems. Commun. Assoc. Inf. Syst. (2018).
2. Bakos, J.Y.: Interorganizational information systems: strategic opportunities for competition and cooperation. Sloan School of Management, Cambridge, MA (1987).
3. Barrett, S., Konsynski, B.: Inter-Organization Information Sharing Systems. MIS Q. 6, 4, 93–105 (1982).
4. Bensaou, M., Venkatraman, N.: Configurations of Interorganizational Relationships: A Comparison Between U.S. and Japanese Automakers, (1995).
5. Blaikie, N.: Designing Social Research. Polity Press, Cambridge (2000).
6. Boonstra, A., de Vries, J.: Analyzing inter-organizational systems from a power and interest perspective. Int. J. Inf. Manage. 25, 6, 485–501 (2005).

7. Chatterjee, D., Ravichandran, T.: Governance of Interorganizational Information Systems: A Resource Dependence Perspective. Inf. Syst. Res. 24, 2, 261–278 (2013).
8. Chatterjee, D., Ravichandran, T.: Inter-organizational Information Systems Research: A Critical Review and an Integrative Framework. In: 37th Hawaii Internatioal Conference on System Sciences. pp. 1–10 (2004).
9. Choudhury, V.: Strategic Choices in the Development of Interorganizational Information Systems. Inf. Syst. Res. 8, 1, 1–24 (1997).
10. de Corbiere, F., Rowe, F.: Understanding the Diversity of Interconnections between IS: Towards a New Typology of IOS. In: Proceedings of 19th European Conference on Information Systems (ECIS 2010). pp. 1–13 (2010).
11. Gregor, S., Johnston, R.B.: Developing an Understanding of Interorganizational Systems: Arguments for Multi Level Analysis and Structuration Theory. In: Proceedings of the European Conference on Information Systems. pp. 1–9 (2000).
12. Hong, I.B.: A new framework for interorganizational systems based on the linkage of participants' roles. Inf. {&} Manag. 39, 4, 261–270 (2002).
13. Klein, S.: Interorganisationssysteme und Unternehmensnetzwerke: Wechselwirkungen Zwischen Organisatorischer und Informationstechnischer Entwicklung. Deutscher Universitäts Verlag, Wiesbaden (1996).
14. Li, J. et al.: A strategic analysis of inter organizational information sharing. Decis. Support Syst. 42, 1, 251–266 (2006).
15. Li, S., Lin, B.: Accessing information sharing and information quality in supply chain management. Decis. Support Syst. 42, 3, 1641–1656 (2006).
16. Lyytinen, K., Damsgaard, J.: Inter-organizational information systems adoption--a configuration analysis approach. Eur. J. Inf. Syst. 20, 496–509 (2011).
17. Meyer, A.D. et al.: Configurational approaches to organizational analysis. Acad. Manag. J. 1175–1195 (1993).
18. Miller, D., Friesen, P.H.: Organizations: A Quantum View. Prentice-Hall, Englewood Cliffs, NJ (1984).
19. Misoch, S.: Qualitative Interviews. De Gruyter Verlag, Berlin (2015).
20. Patton, M.Q.: Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Sage Publications, Thousand Oaks, CA. (2015).
21. Premkumar, G., Ramamurthy, K.: The Role of Interorganizational and Organizational Factors on the Decision Mode for Adoption of Interorganizational Systems*. Decis. Sci. 26, 3, 303–336 (1995).
22. Rodon, J., Sese, F.: Structural Configurations: An Explanation for IOIS Implementation Failure. In: 16th European Conference on Information Systems. pp. 974–984 (2008).
23. Saldaña, J.: The coding manual for qualitative researchers. Sage Publications, London, Thousand Oaks, New Delhi (2009).
24. Sandler, T., Tschirhart, J.: Club Theory: Thirty Years Later. Public Choice. 93, 3/4, 335–355 (1997).
25. Schubert, P.: Integration Scenarios for Business Collaboration. In: AMCIS 2008 Proceedings. p. Paper 258 (2008).
26. Siebert, H.: Ökonomische Analyse von Unternehmensnetzwerken. In: Management von Netzwerkorganisationen. (2010).

Novices' Quality Perceptions and the Acceptance of Process Modeling Grammars

A Trans-disciplinary Quality Approach

Andrea Fürst-Graßl[1]

[1] University of Passau, Chair of Business Information Systems, Passau, Germany
andrea.fuerst@uni-passau.de

Abstract. As Process Modeling Grammars provide a means to visualize and communicate complex business processes, it is crucial to convince novices to adopt them for every-day business. As their drivers of acceptance are widely unknown, my study develops a trans-disciplinary quality approach to investigate how quality perceptions affect novices' adoption intentions. The survey data were analyzed using PLS-SEM. The main result of my study is that the identified quality dimensions are interrelated and differ in their impact on adoption intentions. This provides a 'new', coherent view on quality perceptions of modelling grammars and deeper insights into how they affect behavioral intentions.

Keywords: Business Process Management, Modeling Grammar, Quality Perception, Adoption Behavior, Technology Acceptance

1 Introduction

The increasing complexity and digitalization of business processes requires depicting relevant process information in a clear and transparent manner. Process modeling provides a proper means to visualize, communicate, and evaluate complex business processes [1, 2]. As modeling grammars provide the conceptual base for process modeling by defining a set of graphical constructs and rules for their combination, a standardized modeling grammar is an indispensable prerequisite to integrate process modeling in ever-day business [3, 4]. It enables the use of modeling software to generate process models and to develop a shared understanding of their informational content.

To gain this shared understanding among all employees, the modelers of process models as well as their recipients must have sufficient knowledge about the applied modeling grammar and the willingness to use it in their daily routines [5]. Especially process modeling novices must be encouraged to learn and voluntarily adopt a commonly used modeling grammar. Therefore, it is essential to gain knowledge about the key-drivers of their adoption intentions to foster their acceptance behavior.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Previous research, however, provides only minimal insights into the behavioral mechanisms underlying novices' initial adoption intentions in the context of modeling grammars. Therefore, these relevant key drivers are widely unknown. Prior publications, however, indicate, that quality perceptions are likely to influence the acceptance of a modeling grammar [e.g., 2, 3, 6]. Therefore, this study aims at empirically investigating if and how quality perceptions influence novices' adoption behavior.

The proposed research model develops a two-level quality approach. It builds on the observation, that process modeling novices usually experience a modeling grammar on two levels during a training period: They are taught (1) the language specification including the provided constructs and their rules of interaction and (2) its concrete usage in building process diagrams. Therefore, the quality approach in this study separates between quality perceptions on the *Language Level* and quality perceptions on the *Diagram Level* and poses two main research questions:

1. Do quality perceptions influence novice users' initial acceptance of a modeling grammar?
2. And do the different perceptional levels differ in their impact?

To answer these questions, this research applies a transdisciplinary approach. It aims at taking advantage of the fact that the influence of quality perceptions on adoption intentions is relatively well-investigated in the context of software systems – but poorly investigated in the context of modeling grammars. Therefore, this study identifies conceptual parallels between both research contexts and to transfer valuable insight from the one to the other to gain a well-founded research model as a base of the subsequent empirical investigation.

2 Theoretical Background

2.1 Previous Research on Technology Acceptance

The widespread Technology Acceptance Model (TAM) was originally developed by Davis [7] to explain a user's intention to initially accept a certain technology. Its core statement is, that users intend to accept a certain technology based on their perceptions of its Ease of Use (PEOU) and Usefulness (PU). In recent years, the TAM has been widely applied in the broader context of information systems [8 for an overview]. With regard to process modeling grammars, the publications of Recker [9, 10] showed a good applicability of the TAM in this specific context as well.

Numerous studies on information systems also indicated that quality perceptions may affect the TAM constructs PU and PEOU. Recker et al. [3] were the first who transferred this idea to the context of modeling grammars: They tested a research model combining a quality perspective (focusing on specific ontological deficiencies) with the TAM to investigate experienced users' continuance decisions. Based on their results, it can be presumed that a relationship between quality perceptions and the TAM also exists in the context of modeling grammars. Therefore, my study adopts this approach and builds on merging the TAM with a quality perspective specifically tailored to novice users' quality perceptions.

2.2 The Influence of 'Quality'

Quality-Related Research on Information Systems. In the context of information systems quality perceptions were identified as a key driver of usage intentions and success [11–13]. A clear definition of 'quality', however, is difficult. As quality perceptions are context dependent 'quality' can be interpreted from various perspectives [14–16]. Previous research on quality-driven adoption behavior mainly focused on 'quality' as meeting customer expectations. Particularly the conceptualization originally developed by DeLone & McLean [12] evolved into a de-facto standard.

The original DeLone & McLean model distinguishes between the two dimensions *System Quality* and *Information Quality* which were integrated, adapted, supplemented, and refined in a variety of subsequent investigations [e.g., 12, 14]. *System Quality* reflects the technical component of an information system (e.g., its features and functions) whereas *Information Quality* captures quality perceptions of its informational output like understandability and applicability. Both dimensions were found to have a significant influence on users' intentions to use an information system [12, 13].

Quality-Related Research on Modeling Grammars. In the context of process modeling grammars various researchers consider 'quality' a critical driver of acceptance, too, and continually called for more empirical research in this context [e.g., 2, 3, 6, 17]. However, a clear, commonly used and empirically proven conceptualization of quality in this context is still lacking. Existing quality approaches differ in their conceptualization of quality emphasizing different evaluation criteria, different contexts of usage, and different objectives. Well-regarded publications in this context are for example the SEQUAL-framework focusing on semiotic aspects [e.g., 18, 19], the subsequent publication of Krogstie [5] focusing on model based software development, the CD-framework [20] emphasizing cognitive aspects, Moody's 'Physics of Notations' [2] as a guideline for the design and improvement of modeling grammars, the Bunge-Wand-Weber (BWW) Model [4, 21, 22] focusing on ontological aspects, and the subsequent investigation of Recker et al. [3], focusing on user perceptions of ontological deficiencies.

All of these quality-approaches provide valuable insights into the meaning of 'quality' in the context of modeling grammars. However, due to their specific contexts or missing empirical foundations, none of these approaches provides a proper base for my research. Instead, it seems necessary to merge their core-findings to develop a perception-oriented quality conceptualization focusing on 'quality' as meeting novice users' expectations.

Conceptual Parallels between both Research Contexts. To gain such a conceptualization, it seems appropriate to consider the transferability of findings from the well-investigated field of information systems to the specific context of modeling grammars based on two major parallels:

First, both research fields basically agree that perception-oriented quality approaches are more appropriate to investigate users' acceptance behavior than detailed checklists of objective evaluation criteria [e.g., 3, 11, 23]. Whereas such perception-oriented approaches are already established in the field of information systems, existing quality approaches in the context of modeling grammars predominantly base on detailed and objective evaluation criteria. Moody's "Physics of Notations", for example, provides nine design principles including 25 single criteria [2]. Such detailed evaluation-checklists, however, are not appropriate to investigate novices' adoption behavior as particularly novices were found to develop rather general perceptions including less attributes than knowledgeable individuals [24].

Second, in both research fields, a quality approach including perceptions on different levels seems most appropriate. Quality-related publications in the context of modeling grammars can be divided into two major research streams: Some publications are concerned with quality aspects on the *Diagram Level* whereas others try to make improvements on the *Language Level* [6 for an overview]. Nevertheless, existing quality conceptualizations in the context of modeling grammars do not clearly distinguish between those two perception levels. In contrast, previous research on information systems applies such a two-level approach and clearly separates between S*ystem Quality* and *Information Quality* [12].

This two-dimensional conceptualization for information systems seems well transferable to the context of modeling grammars: As *System Quality* captures the ability of a software system to provide a sound technological base including relevant features and functions [12–14], quality-related publications on the *Language Level* focus on providing a sound constructional base, including ontological, syntactical and semantical aspects of the provided constructs [e.g., 2, 5]. As *Information Quality* is described "in terms of outputs that are useful for business users, relevant for decision making and easy to understand" [14], quality-related publications on the *Diagram Level* emphasize essentially identical issues like applicability, cognitive effectiveness and a clear understandability of the derived process models [e.g., 2, 6, 23, 25, 26]. In both contexts, the quality of the informational output is, to some extent, restricted by the quality of technical or conceptual base.

3 Research Model and Hypotheses Development

3.1 Effects of Quality Perceptions on the Language Level

Quality perceptions on the *Language Level* reflect the evaluation of a modeling grammars specification. As the *Language Level* aims at providing a proper constructional base to generate process models, it corresponds to *System Quality* which aims at providing a proper technical base to generate an informational output.

Gorla et al. [14] suggests to further group the relevant attributes for *System Quality* into the relevant subcategories *System Flexibility* and *System Sophistication*. Previous quality-related research on modeling grammars indicates a similar dichotomy: On the one hand, a wide range and proper definition of the provided constructs – often referred to as *Expressive Power* – is considered fundamentally important [e.g., 2, 3, 28]. On the other hand, previous research claims for a certain level of *Grammar Flexibility* to reduce complexity and use it on different levels of experience [28–30]. Consequently, my study distinguishes between the two constructs *Expressive Power* and *Grammar Flexibility* at the *Language Level*.

Expressive Power refers to the question if a modeling grammar is able to provide constructs to express relevant information completely and concisely [2, 3, 31]. For that aim, especially construct overload and construct deficit must be prevented. Construct overload occurs in situations in which the provided language elements "appear to have multiple real-world meanings and, thus, can be used to describe various real-world phenomena" [31]. This may cause confusion and involves the threat of ambiguity and misinterpretation. Construct deficit occurs when there is no notational element corresponding to a particular real world issue [21]. This causes the problem that relevant facts cannot be expressed by the provided constructional elements. As the Expressive Power of a modeling grammar, therefore, determines its fundamental applicability, I hypothesize

H1a) Expressive Power is positively associated with Perceived Usefulness.

On the other hand a high level of Expressive Power accompanied with a wide range of modeling vocabulary reduces the intuitiveness of a modeling grammar, increases complexity and makes it difficult to learn and handle [28]. Therefore, Expressive Power is supposed to negatively affect its ease of use [28]. I hypothesize

H1b) Expressive Power is negatively associated with Perceived Ease of Use.

Grammar Flexibility addresses the question, if a modeling grammar's specification includes possibilities to reduce it to subsets of core-constructs. This question is of high practical relevance: A study by Sedick & Seymour [28], for example, showed that all surveyed organizations tried to simplify process modeling grammars depending on their individually required level of detail. Zur Muehlen & Recker [29] found out that in every-day business usually a core set of constructs is used and additional constructs are included where necessary. Consequently, a modeling grammar's specification should include a clear distinction between core-sets and expansion sets of constructs (as e.g., BPMN2.0). This may reduce the aforementioned negative effects of increasing complexity to foster an easy usage. Thus, I hypothesize

H2) Grammar Flexibility is positively associated with Perceived Ease of Use.

3.2 Effects of Quality Perceptions on the Diagram Level

Each novice user has a certain set of process models in mind, after finishing a process modeling training. The *Diagram Level* reflects his quality perceptions of these diagrams – as the informational output of a modeling grammar – and, therefore, corresponds to *Information Quality*.

Previous research on information systems suggests breaking *Information Quality* down into content-related and formal evaluation criteria [14]. Content-related quality criteria capture the usefulness and applicability of the provided information whereas the formal criteria correspond to its appearance and understandability [14, 32, 33]. Previous research on process modeling analogously emphasizes a distinction between ‘content’ and ‘format’ with regard to diagrammatic representations [2, 26, 34]. Therefore, my research adopts this dichotomy and distinguishes between *Formal Capability* and *Content Capability* on the *Diagram Level*.

Formal Capability emphasizes the striven goal of the ‘cognitive effectiveness’ of the formal representation of process information in process diagrams [2, 26, 35]. “Cognitive effectiveness is defined as the speed, ease and accuracy with which a representation can be processed by the human mind” [2]. The formal capability of the resulting process diagrams, therefore, reflects a modeling grammar’s ability to visualize processes in a clearly structured and understandable way. BPMN, for example, provides structuring elements like pools and lanes to display interaction, whereas EPC diagrams lack a similar structuring. Following Johansson et al. [36], this leads to a good evaluation of BPMN-models with regard to structure and a rather bad evaluation of EPCs in this context. Consequently, a novice user will probably perceive the underlying modeling grammar as useful and easy to use, if the resulting diagrams appear to foster his efficient information processing. I hypothesize

H3a) Formal Capability is positively associated with Perceived Usefulness.

H3b) Formal Capability is positively associated with Perceived Ease of Use

Content Capability captures the functional perspective of the informational output of a modeling grammar. A process modeling grammar can only be perceived as useful if it provides the ability to build process models for various purposes [2, 3, 18, 23, 27]. If the resulting diagrams seem to be useful in every-day business, to facilitate decision making, and to provide a proper base for communication, the underlying modeling grammar will probably be perceived as useful. Thus, I hypothesize

H4) Content Capability is positively associated with Perceived Usefulness

3.3 Dependencies between the Language Level and the Diagram Level

As a process modeling grammar is not an end in itself but serves the sole purpose of building process models, there may be interdependencies between the quality-constructs on the two levels:

Expressive Power. Though a well-designed modeling grammar does not automatically lead to well-designed process models, a poorly designed modeling grammar makes it impossible to design high-quality diagrams [5]. Therefore, a wide range and proper definition of modeling constructs is a necessary prerequisite for deriving organizational benefits from process modeling [3]. This is especially important to meet the various purposes of process modelling [28]. A lack of well-defined constructional elements, consequently, restricts the applicability of the resulting process diagrams. Therefore, I hypothesize

H5a) Expressive Power is positively associated with Content Capability.

A proper supply of constructional elements affects the *Formal Capability* of the resulting process models as well. Construct overload on the *Language Level* may result in ambiguous process models which include constructs with multiple real-world meanings. This requires users to bring in external knowledge to understand the proper meaning of a construct in a certain context [3, 31] and, therefore, diminishes the understandability of a diagram. Additionally, a modeling grammar may provide – on the *Language Level* – constructs that help to structure the derived diagrams to foster a clear structure and to prevent cognitive overload [2]. Thus, I hypothesize

H5b) Expressive Power is positively associated with Formal Capability.

Grammar Flexibility. A flexible specification with defined subset of constructs may also foster the applicability of the resulting process diagrams. As process models are intended to provide a base for effective communication it "is desirable that a Business Process Model can be understood by the various stakeholders involved in an as straightforward manner as possible" [37]. These process stakeholders may differ with regard to their educational level and modeling experience. Therefore, a subset of core-constructs is helpful for an effective communication between stakeholders on different levels of experience.

H6a) Grammar Flexibility is positively associated with Content Capability.

The reduction to a core-set of constructs may foster the cognitive effectiveness of the resulting diagrams as well. Following Moody [35] the most common mistake in modeling practice is the request to show too much information on a single diagram. The resulting complexity rather impedes than enables effective communication [35]. Therefore, company-wide agreements to only apply a defined subset of constructs may foster the appearance and understandability of the resulting diagrams. Thus, I hypothesize

H6b) Grammar Flexibility is positively associated with Formal Capability.

3.4 The Basic TAM-Hypotheses

The well-established TAM hypotheses have already been tested and verified in numerous TAM-studies [8 for an overview]. Following Recker [9, 10], they turned out to hold in the context of process modeling grammars as well. Thus, I hypothesize:

H7a) Perceived Ease of use is positively associated with Perceived Usefulness

H7b) Perceived Usefulness is positively associated with Intention to Use

4 Research Methodology

4.1 Study Design and Data Collection

This study applies an experimental survey approach, which is a common method to investigate behavioral intentions. Data were collected using a survey of students from a German University of Applied Sciences in January 2017.

In preparation of the survey, all participants were trained in the use of two process modeling grammars (BPMN2.0 and EPCs) during winter term 2016/2017. The students were taught the specification of both modeling grammars (including their notational elements) as well as the concrete use of each grammar for the creation of process models. All of the students were provided the same training documents and the same exercises. Both modeling grammars were taught to a similar extent.

To train each student in two different modeling grammars seemed especially important, as perception development processes are comparative by nature [24].

BPMN2.0 and EPCs were chosen – analogously to the studies by Recker [9, 10] – for two main reasons: on the one hand, both are well-known modeling grammars of high practical relevance focusing on the visualization of information. On the other hand, the two grammars show enough differences with regard to their expressive power and diagram appearance to capture different effects on adoption intentions.

Each participant was asked to answer the questions in the questionnaire for each of the two modeling grammars. As all of the participants were students of a German-speaking class, the questions were provided in German language as well. To ensure content-equivalence between the German and English version of the measurement items, the translation procedure recommended in Brislin [38] was applied.

I received 44 completed and usable questionnaires, each including assessments of the two different process modeling grammars. This resulted in 44*2=88 total observations for further analysis. Among the 44 participants, 20.5% were female, 79,5% were male. The average age was 22.7 years. The participants were all students of an Information Systems Bachelor Degree Program (100%).

4.2 Construct Measurement

Due to a lack of appropriate measures in Process Modelling Research, the measurement items for *Expressive Power, Grammar Flexibility, Formal Capability* and *Content Capability* were derived from the *System Quality* and *Information Quality* measures in the study of Gorla [14]. These measures represent a well-founded synthesis of quality-related measures from various Information Systems studies (see [14] for an overview). As these measures were designed for the evaluation of Information Systems, they had to be reformulated and adapted to the specific context of process modelling. These adaptions were discussed with several experienced researchers to ensure content validity and understandability of the resulting measures.

The measures for the TAM constructs are based on the publications of Recker [3, 9, 10, 39], Moore and Benbasat [40], and Venkatesh and Davis [41]. All constructs were measured reflectively on a 7-point Likert Scale (1=”fully disagree” to 7=”fully agree”).

Table 1. Measurement Items of the Applied Constructs

Expressive Power (EXP)
The process modeling grammar provides notational elements to…
…capture information accurately
…capture information completely
…capture information concisely
Grammar Flexibility (FLEX)
The process modeling grammar…
…can be reduced to a set of individually useful features and functions
…can be handled by all levels of users
Formal Capability (FORM)
The resulting process depiction…

…has a good appearance …has a clear structure …is clearly understandable
Content Capability (CC)
The created process models … …are useful for the professional work …facilitate decision-making …provide a proper foundation for communication
Perceived Ease of Use (PEOU)
Overall, I believe, that using this process modeling grammar is easy. Modeling processes in the intended way is easy with this modeling grammar I find creating process models using this modeling grammar is easy.
Perceived Usefulness (PU)
Overall, I find this modeling grammar useful for process modeling. I consider this modeling grammar appropriate for process modeling. I find this modeling grammar useful for meeting my process modeling objectives
Intention to Use (INTU)
If I have access to this modeling grammar, I intend to use it for process modeling. My intention is to use this modeling grammar for process modeling. I prefer to use this modeling grammar instead of using another process modeling grammar.

5 Measurement Validation and Hypotheses Testing

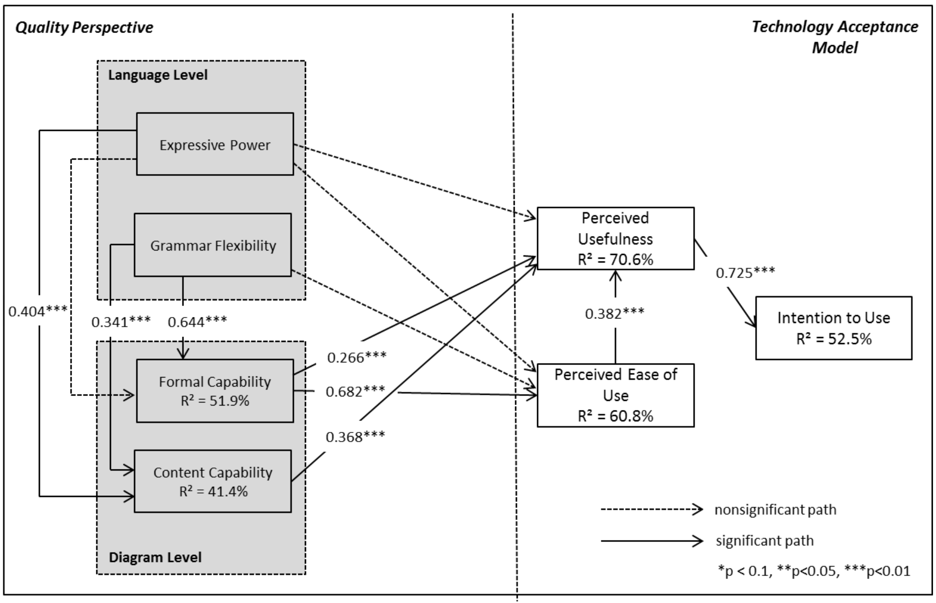


Figure 1. Results of the PLS-estimation of the research model

As the quality-driven adoption of modeling grammars is poorly investigated so far, applying the PLS-SEM method seems appropriate to meet the exploratory character of this study. This method is particularly suitable to test theories in early stages, as it makes fewer demands on data distributions and sample sizes compared to covariance-based approaches [42, 43]. Based on the research of Cohen [44], Hair et al. [43] recommend a minimum of 58 observations for a respective research model with maximally four arrows pointing at a construct. As 88 observations easily exceed this recommendation, the sample size should be sufficient for a sound data analysis.

Table 2. Cronbach's Alpha, CR, AVE and HTMT

Constr.	*Cr. α*	*CR*	*AVE*	*EXP*	*INTU*	*PEOU*	*PU*	*FORM*	*CC*
EXP	0.880	0.925	0.805						
INTU	0.928	0.954	0.874	**0.355**					
PEOU	0.923	0.951	0.866	**0.468**	**0.701**				
PU	0.927	0.954	0.873	**0.473**	**0.776**	**0.802**			
FORM	0.884	0.929	0.813	**0.507**	**0.677**	**0.854**	**0.823**		
CC	0.829	0.897	0.745	**0.652**	**0.640**	**0.551**	**0.762**	**0.669**	
FLEX	0.744	0.886	0.796	**0.602**	**0.504**	**0.726**	**0.646**	**0.874**	**0.669**

The **evaluation of the measurement model** followed the established evaluation criteria recommended in Hair et al. [43]. With regard to **Internal Consistency Reliability** all constructs meet the established quality criteria recommended in Bagozzi & Yi [45] and Hair et al. [43] (Cronbach's $\alpha > 0.7$ and $CR > 0.7$). With all outer loadings exceeding the recommended threshold of 0.7 [43], **Indicator Reliability** is also given. As all AVE values are higher than the recommended threshold of 0.5 [43], **Convergent Validity** is also fulfilled. The **Constructs' Discriminant Validity** was evaluated by applying the recommended HTMT approach [43, 46]. Henseler et al. [46] propose two types of limits for HTMT-values: a strict limit of 0.85 and a more permissive limit of 0.9. 19 of 21 HTMT-values in this study meet the strict threshold of 0.85; only two exceed this limit minimally (0.874 and 0.854) but are below the limit of 0.9 (see Tabel 2, bolded values).

Hair et al. [47] propose to **validate the structural model** as follows: To avoid critical levels of **collinearity** among the predictor constructs, computing the VIF values for all predictor variables is recommended. All VIF values turned out to be far below the established limit of 5 [47]. The amount of **variance explained** (R^2) was considerable high exceeding the level of 50% for all of the endogenous TAM constructs and exceeding 40% for all of the endogenous constructs on the Diagram Level (see figure 1). The **cross-validated redundancy value Q^2** [48, 49] was >0 for each of the endogenous constructs (0.428 for INTU, 0.481 for PEOU and 0.543 for PU, 0.383 for FORM, and 0.278 for CC), which indicates predictive relevance [50].

The **path coefficients** of the proposed hypotheses were computed using the PLS-SEM algorithm implemented in SmartPLS [51] (see Figure 1). To test the significance of each path, the corresponding **t-values** were computed applying the PLS SEM Bootstrapping Routine in Smart PLS with 5000 subsamples and a two-tailed test. All hypothesized relationships between the quality constructs on the *Diagram Level* and the TAM constructs were supported and turned out to be significant at a 1% level. Significant paths between the *Language and the Diagram Level* could be identified as well: The hypothesized relationships between *FLEX* and FORM, between *FLEX* and *CC*, and between *EXP* and *CC* were found to be significant at a 1% level. The assumed relationship between *EXP* and *CC* was not supported. Interestingly, none of the hypothesized direct relationships between the constructs on the *Language Level* and the TAM constructs was significant. This surprising result will be discussed in detail in the subsequent section.

6 Results, Conclusions and Future Research

This study posed two research questions in the introduction. With regard to the first question, a clear link between quality-perceptions and novices' usage intentions could be found and the identified quality dimensions could explain a considerable portion of the variance of the TAM-constructs. With regard to the second question, the identified quality dimensions turned out to influence the TAM-Constructs in different ways. Whereas the perceptions on the *Diagram Level* directly affected the students' intention to use a modeling grammar, the perceptions on the *Language Level* affected their usage intentions only indirectly via the *Diagram Level*. Consequently, the *Diagram Level* can be interpreted as a kind of perceptional mediator between the *Language Level* and the resulting acceptance intentions. This somewhat surprising result can be explained by the fact that information in diagram form can be processes and remembered better than ordinary language [2 for an overview]. Information about a modeling grammar in diagram form, consequently, is likely to have a stronger and instant influence on the subsequent acceptance behavior than the rather abstract grammar specification. Quality perceptions about the specification, however, are not irrelevant, as they do influence the users' perceptions on the Diagram Level.

In summary, the main results of my investigation are that (1) considerable parallels between the 'quality' of information systems and modeling grammars can be found, (2) that a coherent, perception-oriented approach is appropriate to capture novice users' quality perceptions, (3) that these quality perceptions do influence the initial acceptance of a modeling grammar, and (4) that the identified perceptional levels differ in their cause-effect relationships. They contribute to theory and practice:

From a **theoretical perspective**, my study introduces a 'new' view on quality perceptions in the context of modeling grammars. It clearly indicates that perceptions on the Language Level cannot be investigated separately from perceptions on the Diagram Level, as one depends on the other. By combining the Language and the Diagram Perspective, it merges the core-subjects of two wide research streams into a single quality model. This may provide a base to better understand and further investigate open questions from prior research. Recker et al. [3], for example, received mixed and inconsistent results whether certain perceptions of ontological deficiencies directly affect PEOU and PU. Building on my results, it seems possible that perceptions of some ontological weaknesses influence subsequent perceptions of PEOU and PU only indirectly via perceptions on a Diagram Level.

From a **managerial perspective** my results may help (1) to choose a modeling grammar that will probably be voluntarily accepted among all employees and (2) to design proper training strategies:

With regard to selection decisions, my study showed that *Expressive Power* and *Grammar Flexibility* both positively affect users' quality perceptions of the resulting diagrams and subsequently their adoption intentions. This indicates that it is crucial to select a modeling grammar that provides – on the one hand – a wide and well-defined supply of constructs and – on the other hand – is kept flexible enough to work on individually required subsets.

Knowing about the drivers of novices' acceptance intentions (including their cause-effect relationships) may help to develop appropriate training strategies as well: It seems reasonable to first introduce a modeling grammar on the Language Level and to teach the concrete usage for building process diagrams in a second step. If modeling novices were first shown the provided constructs as well as recommended subsets, this knowledge is likely to influence their subsequent perception of the resulting diagrams – which was found to directly influence their adoption intentions.

As this study has a few **limitations** as well it encourages further research in the following areas: First, the study is an exploratory approach to provide a first insight into the relevance of quality perceptions for novice users' adoption behavior. It was based on novices' perceptions of only two modeling grammars. Future research should extend this study on bigger sample sizes and additional modeling grammars.

Secondly, the focus of this study was on novice users. The level of individual experience may, however, influence individual quality perception development processes. Therefore, the study needs to be repeated with a respective sample of more experienced users.

References

1. Green, P.F., Rosemann, M.: Perceived ontological weaknesses of process modeling techniques: Further evidence. ECIS 2002 Proceedings. 104 (2002).
2. Moody, D.: The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on Software Engineering. 35, 756–779 (2009).
3. Recker, J., Rosemann, M., Green, P., Indulska, M.: Do ontological deficiencies in modeling grammars matter? MIS Quarterly. 57–79 (2011).
4. Wand, Y., Weber, R.: Research commentary: information systems and conceptual modeling—a research agenda. Information systems research. 13, 363–376 (2002).
5. Krogstie, J.: Model-based development and evolution of information systems: A Quality Approach. Springer Science & Business Media, London (2012).
6. Genon, N., Heymans, P., Amyot, D.: Analysing the cognitive effectiveness of the BPMN 2.0 visual notation. In: International Conference on Software Language Engineering. pp. 377–396. Springer (2010).
7. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly. 319–340 (1989).

8. Legris, P., Ingham, J., Collerette, P.: Why do people use information technology? A critical review of the technology acceptance model. Information & management. 40, 191–204 (2003).
9. Recker, J.: Why do we keep using a process modelling technique?, 18th Australasian Conference on Information Systems. 49-59 (2007).
10. Recker, J.: Explaining usage of process modeling grammars: Comparing three theoretical models in the study of two grammars. Information & management. 47, 316–324 (2010).
11. Agarwal, R., Prasad, J.: The antecedents and consequents of user perceptions in information technology adoption. Decision support systems. 22, 15–29 (1998).
12. DeLone, W.H., McLean, E.R.: Information systems success: The quest for the dependent variable. Information systems research. 3, 60–95 (1992).
13. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: a ten-year update. Journal of management information systems. 19, 9–30 (2003).
14. Gorla, N., Somers, T.M., Wong, B.: Organizational impact of system quality, information quality, and service quality. The Journal of Strategic Information Systems. 19, 207–228 (2010).
15. Reeves, C.A., Bednar, D.A.: Defining quality: alternatives and implications. Academy of management Review. 19, 419–445 (1994).
16. Swanson, E.B.: Maintaining IS quality. Information and Software Technology. 39, 845–850 (1997).
17. Störrle, H., Fish, A.: Towards an Operationalization of the "Physics of Notations" for the Analysis of Visual Languages. In: International Conference on Model Driven Engineering Languages and Systems. pp. 104–120. Springer (2013).
18. Krogstie, J., Sindre, G., Jørgensen, H.: Process models representing knowledge for action: a revised quality framework. European Journal of Information Systems 15, 91–102 (2006).
19. Lindland, O.I., Sindre, G., Solvberg, A.: Understanding quality in conceptual modeling. IEEE software. 11, 42–49 (1994).
20. Green, T.R., Blandford, A.E., Church, L., Roast, C.R., Clarke, S.: Cognitive dimensions: Achievements, new directions, and open questions. Journal of Visual Languages & Computing. 17, 328–365 (2006).
21. Wand, Y., Weber, R.: An ontological model of an information system. IEEE transactions on software engineering. 16, 1282–1292 (1990).
22. Wand, Y., Weber, R.: On the ontological expressiveness of information systems analysis and design grammars. Information Systems Journal. 3, 217–237 (1993).
23. Maes, A., Poels, G.: Evaluating quality of conceptual modelling scripts based on user perceptions. Data & Knowledge Engineering. 63, 701–724 (2007).
24. Steenkamp, J.-B.E.: Conceptual model of the quality perception process. Journal of Business research. 21, 309–333 (1990).
25. Figl, K.: Comprehension of Procedural Visual Business Process Models. Business and Information Systems Engineering. 59, 41–67 (2017).
26. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cognitive science. 11, 65–100 (1987).
27. Van der Aalst, W.M.: Business process management: a comprehensive survey. ISRN Software Engineering. 2013, (2013).
28. Sedick, F., Seymour, L.F.: BPMN usage: An analysis of influencing factors. Advances in Enterprise Information Systems II. 155 (2012).
29. zur Muehlen, M., Recker, J.: How much language is enough? Theoretical and Practical Use of the Business Process Management Notation. In: Leonard, M., Ballahsène (eds.)

Advanced Information Systems Engineering - CAiSE 2008, Montpellier, France. pp. 465-479. Springer (2008).
30. zur Muehlen, M., Recker, J.: How much language is enough? Theoretical and practical use of the business process modeling notation. In: Bubenko J., Krogstie, J., Pastor, O., Pernici, B., Rolland, C., Solvberg, A. (eds.) Seminal Contributions to Information Systems Engineering. pp. 429–443. Springer, Berlin, Heidelberg (2013).
31. Recker, J., Rosemann, M., Indulska, M., Green, P.: Business process modeling-a comparative analysis. Journal of the Association for Information Systems. 10, 1 (2009).
32. Nelson, R.R., Todd, P.A., Wixom, B.H.: Antecedents of information and system quality: an empirical examination within the context of data warehousing. Journal of management information systems. 21, 199–235 (2005).
33. Wang, R.Y., Strong, D.M.: Beyond accuracy: What data quality means to data consumers. Journal of management information systems. 12, 5–33 (1996).
34. Siau, K.: Informational and computational equivalence in comparing information modeling methods. Journal of database management. 15, 73 (2004).
35. Moody, D.: What makes a good diagram? improving the cognitive effectiveness of diagrams in is development. In: Wojtkowski, W., Wojtkowski, W.G., Zupancic, J., Magyar, G. Knapp, G. (eds.) Advances in information systems development. pp. 481–492. Springer (2007).
36. Johansson, L.-O., Wärja, M., Carlsson, S.: An evaluation of business process model techniques, using Moody's quality criterion for a good diagram. In: CEUR Workshop proceedings. pp. 54–64 (2012).
37. Van der Aalst, W. M. P., Hofstede, T., Weske, M.: Business process management: A survey. In: Proceedings of the 1st International Conference on Business Process Management, volume 2678 of LNCS. Citeseer (2003).
38. Brislin, R.W.: Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology. 1, 185–216 (1970).
39. Recker, J.C., Rosemann, M.: A measurement instrument for process modeling research: development, test and procedural model. Scandinavian Journal of Information Systems. 22, 3–30 (2010).
40. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Information systems research. 2, 192–222 (1991).
41. Venkatesh, V., Davis, F.D.: A model of the antecedents of perceived ease of use: Development and test. Decision sciences. 27, 451–481 (1996).
42. Fornell, C., Bookstein, F.L.: Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing research. 440–452 (1982).
43. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modeling. Sage Publications Ltd., Los Angeles (2016).
44. Cohen, J.: Statistical power analysis for the behavioral sciences . Hilsdale. NJ: Lawrence Earlbaum Associates. 2, (1988).
45. Bagozzi, R.P., Yi, Y.: Specification, evaluation, and interpretation of structural equation models. J. of the Acad. Mark. Sci. 40, 8–34 (2012).
46. Henseler, J., Ringle, C., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science. 43, 115–135 (2015).
47. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modeling. Sage Publications Ltd., Los Angeles (2016).

48. Geisser, S.: The predictive sample reuse method with applications. Journal of the American statistical Association. 70, 320–328 (1975).
49. Stone, M.: Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological). 111–147 (1974).
50. Chin, W.W.: How to write up and report PLS analyses. In: Esposito Vinzi, V., Chin, W.W., Henseler, J., Wang H. (eds.) Handbook of partial least squares. pp. 655–690. Springer, Berlin, Heidelberg (2010).
51. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. Boenningstedt: SmartPLS GmbH (2015).

Entwicklung einer Definition für Social Business Objects (SBO) zur Modellierung von Unternehmensinformationen

Berit Gebel-Sauer[1], Petra Schubert[1]

[1] Universität Koblenz-Landau, FG Betriebliche Anwendungssysteme, Koblenz, Deutschland
{gebelsauer,schubert}@uni-koblenz.de

Abstract. Der folgende Beitrag enthält Ergebnisse aus dem Langzeitforschungsprojekt SoNBO *(Social Network of Business Objects)*, dessen Ziel es ist, ein generisches Konzept für die Integration von Geschäftsinformationen aus verschiedenen Anwendungssystemen zu entwickeln. Nachdem zwei prototypische SoNBO-Anwendungen in einem strukturierten *Design-Science-Research-(DSR)-Prozess* erstellt wurden, wurde in der Folge die Definition des für SoNBO zentralen Begriffs des *„Social Business Objects"* entwickelt und nachvollziehbar dokumentiert. Quelle für die Entwicklung der Definition waren eine strukturierte Literaturanalyse, die Auswertung von Interviews mit Unternehmensvertretern und eine Analyse der zwei bereits existierenden Anwendungen. Den drei Quellen war gemein, dass unter *Business Objects* eine Klassifizierung bzw. Kategorisierung von Unternehmensinformationen für die Speicherung in Informationssystemen verstanden wird. Der SoNBO-Ansatz greift dieses Verständnis auf und macht die Struktur von *Social Business Objects* explizit für den User in Enterprise Social Software sichtbar (im Frontend). Die vorgenommene Schärfung der Terminologie ist ein wichtiger Meilenstein im SoNBO-Langzeitforschungsprojekt und dient als Input für den nächsten DSR-Zyklus.

Keywords: Social Business Object, Geschäftsobjekt, Integration, Information, SoNBO

1 Motivation und wissenschaftliche Einordnung

Betriebliche Anwendungssysteme werden in Unternehmen zur Beschaffung, Verarbeitung, Übertragung, Speicherung und/oder Bereitstellung von Informationen genutzt und sind das Rückgrat der elektronischen Informationsverarbeitung [1]. Sie lassen sich grob in vier Softwaretypen unterteilen: (Typ 1) Betriebliche Kernsysteme, häufig in der Form von *Enterprise-Resource-Planning-Systemen (ERP-Systeme)* (z. B. SAP ERP) und *Customer-Relationship-Management-Systemen (CRM-Systeme)* (z. B. Salesforce), unterstützen die funktionalen Bereiche eines Unternehmens und bestehen aus unterschiedlichen Modulen, wie z. B. Finanzmanagement, Materialmanagement, Personalmanagement, Verkauf/Marketing oder Einkauf [2]. Diese Systeme speichern

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

und verarbeiten unternehmenskritische Informationen in der Form von *Business Objects (BO)* (z. B. Personen oder Produkte), die für die Tätigkeit der Wertschöpfung notwendig sind. (Typ 2) *Enterprise Collaboration Systems (ECS)* (Kollaborationssysteme, z. B. IBM Connections) unterstützen Mitarbeitende in ihrer gemeinsamen Arbeit (z. B. an Dokumenten oder in der Kommunikation) und sind damit komplementär zu den prozessorientierten ERP-Systemen. ECS umfassen Module für die Kommunikation, Kooperation, Koordination und die gemeinsame Arbeit an Content [3]. Sie kombinieren traditionelle Groupware-Komponenten (z. B. E-Mail oder Gruppenkalender) mit Social-Software-Komponenten (z. B. soziale Profile, Wiki, Blog) [4, 5]. (Typ 3) *Enterprise-Content-Management-Systeme (ECMS)* (z. B. Alfresco) stellen Software bereit, um Content wie z. B. Grafiken oder Dokumente zu verwalten oder zu archivieren. Sie bilden das Bindeglied zwischen ERP-Systemen und ECS. (Typ 4) Ein *Business-Process-Management-System (BPMS)* (z. B. Camunda) ist ein Softwaresystem, das Arbeitsabläufe mit Hilfe der zugrundeliegenden Prozessdefinitionen steuert und die Umsetzung von Geschäftsprozessen koordiniert [6]. Im Rahmen der darin durchgeführten Prozessorchestrierung werden Funktionen der drei zuvor genannten Softwaretypen bedarfsweise im Prozessablauf aufgerufen. Neben diesen Haupttypen betrieblicher Informationssysteme, die alle als Standardsoftware am Markt verfügbar sind, findet sich in Unternehmen eine Vielzahl an geschäftsspezifischen Spezialapplikationen (meist in der Form von Individualsoftware), die speziell für dieses Unternehmen entwickelt wurden.

Historisch bedingt sind geschäftsrelevante Informationen auf verschiedene Informationssysteme verteilt, was eine umfassende und konsistente Sicht auf Business Objects erschwert [7]. Daher ist es notwendig geworden, die benötigten Informationen mit der Hilfe von Integrationslösungen aus verschiedenen betriebswirtschaftlichen Anwendungssystemen zu aggregieren, um sie den Mitarbeitenden (idealerweise unter einer einheitlichen Oberfläche) zur Verfügung zu stellen. Für diese Herausforderung gibt es bereits technische Ansätze zur Integration, die in der akademischen Literatur z. B. unter den Stichworten [8] *Enterprise Application Integration* (EAI) [9, 10] oder *Service Oriented Architecture* (SOA) [11, 12] diskutiert werden. Bei diesen existierenden Ansätzen wird in der Konzeptionsphase zunächst der Informationsbedarf der Anwender ermittelt, woraufhin eine Integrationslösung (Softwarekomponente) entwickelt und eingeführt wird [9, 13]. Allerdings ist der mit diesen Lösungen (z. B. SOA) verbundene Aufwand oftmals zu hoch, um eine flexible und effektive Datenintegration in der heutigen dynamischen Unternehmenswelt zu realisieren [14].

Vor diesem Hintergrund wurde **SoNBO** *(Social Network of Business Objects)* entwickelt, ein *graphenbasiertes Konzept* für eine Informationsintegration. In einem SoNBO [15, 16] werden die Geschäftsinformationen der betrieblichen Anwendungssysteme in Klassen kategorisiert (z. B. Mitarbeitender oder Angebot) sowie deren Attribute und die Beziehungen zwischen den Klassen identifiziert. Zu den Klassen gibt es jeweils Instanzen (vgl. Abbildung 1, links). Das auf diese Weise entstehende Netzwerk (auf Klassenebene) wird in einer SoNBO-Anwendung (vgl. Abbildung 1, rechts) konfiguriert und das Frontend wird als Webanwendung in einem ECS integriert.

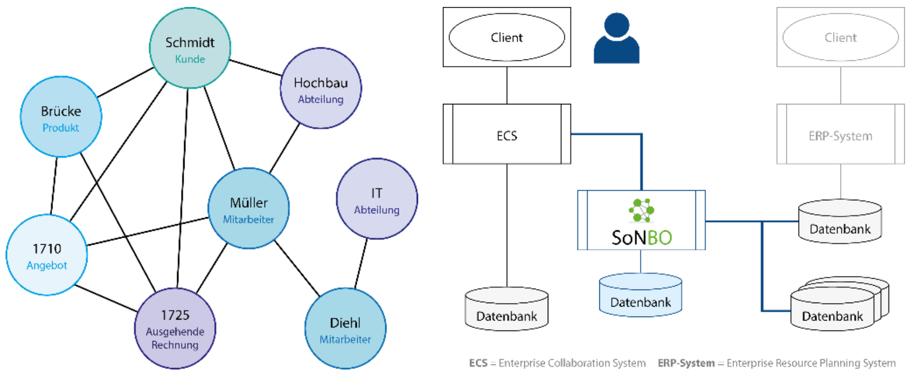


Abbildung 1. SoNBO-Ansatz (verändert nach [15])

Dadurch werden die Informationen der transaktionalen Systeme (ERP-/CRM-System) in das Kollaborationssystem (ECS) integriert. Wenn ein User eine Information zu einer konkreten Instanz (= BO) einer Klasse benötigt, wird dieses BO zur Laufzeit mit Hilfe von Konfigurationseinstellungen aufgerufen (z. B. Angebot *1710* oder Kunde *Schmidt*), die vorher für die zugehörige Klasse festgelegt wurden. Beispielsweise werden im Frontend (vgl. Abbildung 2, Mitte unten) die umliegenden Business Objects angezeigt, sodass der User auf das nächste BO bzw. Knoten navigieren („surfen“) kann.

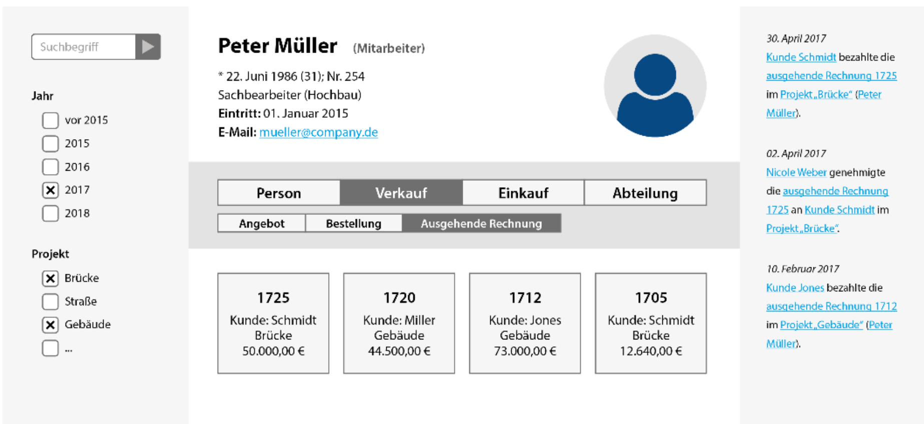


Abbildung 2. Mock-up einer SoNBO-Anwendung

Gewehr et al. [15] beschreiben das Konzept als innovativ für die Informationsintegration, da durch diese Art der Integration die Informationen in den Systemen zur Laufzeit abgerufen werden und somit keine redundante Datenspeicherung erforderlich ist. Außerdem kann dieses Konzept flexibel auf Änderungen im Unternehmen reagieren, indem das Netzwerk durch Konfiguration angepasst werden kann. Im Forschungsbereich Semantic Web gibt es zwei vergleichbare graphenbasierte Integrationsansätze, die das abstrakte Konzept auf die konkrete Anwendung in Unternehmen übertragen: *Linked Enterprise Data* [14] und *Enterprise Knowledge Graph* [17].

Wie der Name *Social Network of Business Objects* (SoNBO) bereits andeutet, spielen die *Business Objects* und daraus abgeleitet die *Social Business Objects (SBO)* in diesem Konzept eine zentrale Rolle. Die Entwickler des Konzepts benutzen bei SoNBO den Begriff Business Object (deutsch: Geschäftsobjekt) bzw. *Social Business Object* für die konkreten Instanzen (Schmidt) einer Klasse (Kunde).

Der SoNBO-Ansatz wurde bereits mit Hilfe von zwei konkreten Anwendungen in der Praxis evaluiert. Die erste SoNBO-Anwendung wurde in einem Unternehmen entwickelt, implementiert und wird seitdem von den Mitarbeitenden aktiv genutzt. Die Evaluationsergebnisse wurden in Form einer Fallstudie dokumentiert [15]. Die (positiven) Ergebnisse aus diesem Projekt gaben den Anstoß zu einer Vertiefung der Forschung im Bereich SoNBO. Das erste Projekt war erfolgreich in einer spezifischen Branche mit einer spezifischen Problemstellung. Die weitere SoNBO-Forschung verfolgt das Ziel der Entwicklung eines generischen Konzepts, das die Übertragbarkeit auf beliebige weitere Unternehmen ermöglicht. Die Herausforderung bei der Implementation einer SoNBO-Applikation besteht u. a. darin, das *unternehmensspezifische Netzwerk* zu identifizieren und dieses dann zu konfigurieren. Für die Evaluation der Übertragbarkeit benötigten die Autoren dieses Papers daher zunächst eine eigene SoNBO-Anwendung, um mit deren Hilfe die notwendigen Schritte zur Konfiguration zu identifizieren und anschließend ggf. zu generalisieren. Während der Entwicklung der Anwendung stellte sich heraus, dass eine eindeutige Definition eines Social Business Objects für das Konzept von zentraler Bedeutung ist, dass aber eine allgemein anerkannte Definition in der Literatur zum jetzigen Zeitpunkt nicht existiert. Fachgebiete wie das Informationsmanagement werden von dieser Schärfung der Begrifflichkeiten profitieren, da durch SoNBO Unternehmensinformationen mittels einer Ontologie in ein Unternehmensnetzwerk (= Graphenstruktur) kategorisiert und somit organisiert werden.

Ziel des Papers ist daher die Entwicklung einer Definition für Social Business Objects. Da diese eng mit Business Objects verbunden sind, ergeben sich die folgenden zwei Forschungsfragen:

1. Status Quo: Welches Verständnis über ein Business Object existiert aktuell in Wissenschaft und Praxis?
2. Definition: Wie kann auf dieser Basis eine Definition für ein *Social* Business Object abgeleitet werden?

2 Forschungsdesign

Wie bereits zuvor erläutert, ist die Entwicklung einer Definition für den Begriff des Social Business Objects Teil eines Langzeitforschungsprojekts. Nachdem eine selbst entwickelte Individualsoftware (im Folgenden referenziert als: „erste“ SoNBO-Anwendung) in einem mittelständischen Unternehmen als innovatives Konzept für eine Informationsintegration von den Forschern identifiziert wurde, wurde diese durch eine Tiefenfallstudie [18, 19] erhoben und veröffentlicht [15]. Um das übergeordnete

Forschungsziel zu erreichen (aus diesem Fallbeispiel eine Übertragbarkeit zu generieren) wurde nach der Methodik *Design Science Research (DSR)* [20] der SoNBO-Explorer (die „zweite" SoNBO-Anwendung) entwickelt, konfiguriert und die Ergebnisse dazu ebenfalls veröffentlicht [16]. Der SoNBO-Explorer integriert Informationen aus einem produktiv genutzten CRM-System einer universitären Forschungsgruppe (GEDYS IntraWare) in deren Kollaborationssystem (IBM Connections). Im Zuge der Entwicklung und Konfiguration stellte sich heraus, dass die verwendete Terminologie (Social Business Object) in der Literatur unterschiedlich genutzt wird und für den SoNBO-Ansatz präzisiert werden muss. Die Entwicklung der Begriffsdefinition folgte der Empfehlung des DSR-Zyklus nach Kuechler & Vaishnavi [20]. Der entstandene Erkenntnisgewinn dient als Input für den nächsten Durchlauf des DSR-Zyklus (hier: Fortsetzung der SoNBO-Forschung mit dem oben angesprochenen Ziel).

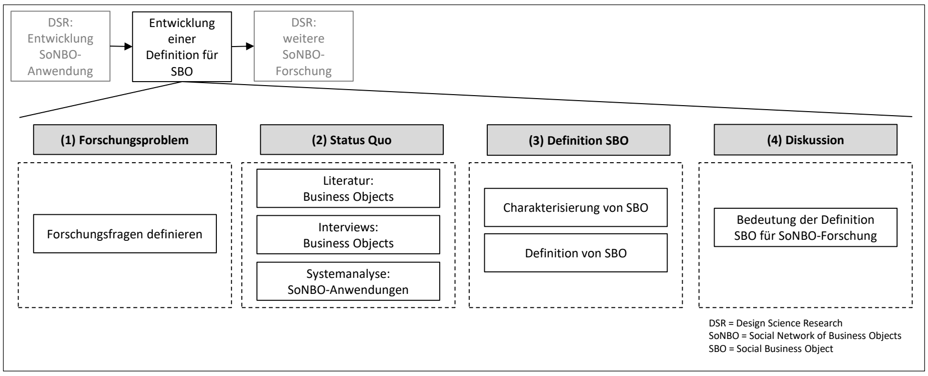


Abbildung 3. Forschungsschritte

Die Forschungsschritte, die zu der Definition über Social Business Objects führen, lassen sich folgendermaßen strukturieren (vgl. Abbildung 3): Nachdem die zweite SoNBO-Anwendung entwickelt und konfiguriert wurde, ergaben sich (1) die Forschungsfragen, die bereits in Kapitel 1 vorgestellt wurden.

Die Etablierung des Status Quo begann mit (2) einer Literaturanalyse [21, 22] zur Verwendung des Begriffs in der *akademischen Literatur*. Dazu wurden die Schlagworte *„Business Object"*, *„Business Objekt"* und *„Geschäftsobjekt"* in google scholar und in den Literaturdatenbanken verwendet, die den Autoren über die Universität zur Verfügung stehen. Für die Auswahl der Literaturquellen wurde keine zeitliche Einschränkung bezüglich des Publikationsdatums vorgenommen. Anhand des Abstracts erfolgte eine Vorauswahl von relevanten Beiträge. Anhand der Literaturverzeichnisse wurden weitere Beiträge identifiziert. Relevante Beiträge wurden systematisch ausgewertet und die Ergebnisse wurden kategorisiert. Die Kategorisierung diente als Ausgangspunkt für die Entwicklung einer Definition in diesem Paper. Da der Fokus des vorliegenden Papers auf der Entwicklung einer Definition zu Social Business Objects liegt, wird die Literaturanalyse in der Form der wesentlichen Ergebnissen vorgestellt, um das grundsätzliche Verständnis aus der Literatur zu vermitteln.

Für die Etablierung eines groben Verständnisses zur Verwendung des Begriffs „Business Object“ *in der Praxis*, wurden sechs ausgewählte Unternehmensvertreter an einem Fachkongress interviewt [23]. Die qualitativen Interviews waren offen und geleitet durch die konkrete Frage „Was verstehen Sie unter einem Business Object?“. Den Befragten war gemeinsam, dass sie alle einen technischen Hintergrund (IT-Abteilung) hatten. Außerdem benutzten die Befragten betriebliche Anwendungssysteme (auch ERP-Systeme) intensiv im Berufsalltag. Allerdings waren diese nicht an der Entwicklung von ERP-Systemen beteiligt. Die Antworten wurden im Anschluss ebenfalls analysiert und kategorisiert. Im nächsten Schritt fand eine Analyse der Software der beiden existierenden SoNBO-Anwendungen statt, um das darin enthaltene Verständnis zu Business Objects zu identifizieren.

Die folgenden Abschnitte beschreiben die Entwicklung der Definition (3). Den Einstieg bildet eine Charakterisierung eines Social Business Objects mit Beispielen. Anschließend wird eine Definition für ein Social Business Object mit Hilfe der drei Ergebnisse des Status Quo (Literatur, Praxis, Systemanalyse) präsentiert. Abschließend wird (4) die Bedeutung der Definition für die weitere SoNBO-Forschung erläutert.

3 Business Objects

Der Begriff *Business Object* wird sowohl in der wissenschaftlichen Literatur als auch in der Praxis verwendet. Im Folgenden werden die Ergebnisse der Literaturanalyse, die Auswertung der Interviews mit den Praxisvertretern und anschließend das Ergebnis der Systemanalyse der beiden existierenden SoNBO-Anwendungen und das darin implizit enthaltene Verständnis zu einem Business Object präsentiert.

3.1 Ergebnis der Literaturanalyse

Die identifizierten Beiträge in der wissenschaftlichen Literatur zu Business Objects lassen sich in *drei Bereiche* kategorisieren, die in Tabelle 1 dargestellt sind: (1) Softwareentwicklung, (2) Softwareeinführung und (3) Softwareorchestrierung. Zu jedem Bereich ist ein Zitat zu dem Begriff Business Object gelistet, das charakterisierend für das Verständnis in dem jeweiligen Fachgebiet ist.

Business Objects sind für die *Informati*k und insbesondere für die *Softwareentwicklung* von Bedeutung. Mitte der 1990er Jahre führte die Business Object Management Special Interest Group (BOMSIG) (später: BODTF = Business Object Domain Task Force) als Einheit der Object Management Group (OMG) das Business Object-Konzept ein. Das Ziel bestand darin, softwaretechnische und betriebswirtschaftliche Modelle stärker miteinander zu verbinden sowie auf der Basis von CORBA (= Common Object Request Broker Architecture) ein Framework zu spezifizieren, das Softwareentwickler unterstützen sollte, komplexe Anwendungssysteme unter der Verwendung von Business Objects zu entwickeln [24]. SAP definierte beispielsweise ein Business Framework für eine komponentenbasierte Architektur, mit dessen Hilfe die Softwarekomponenten von verschiedenen Herstellern interoperieren können [24]. Das SAP-Datenmodell beschreibt die zugrundeliegende betriebswirtschaftliche Struktur in über 4000 Entitätstypen [25].

Es gibt zwei weitere Frameworks, die beschreiben, wie Business Objects strukturiert sein müssen, damit sie zu komplexen betrieblichen Anwendungssystemen zusammengefügt werden können: *OMG Business Object Facility* und *San Francisco Framework*. Unter Business Objects werden hier ausführbare Softwarekomponenten verstanden, die die Struktur und das Verhalten von betriebswirtschaftlich relevanten Einheiten kapseln [26]. Es handelt sich dabei um die Repräsentation eines Gegenstandes aus dem realen Geschäftsleben und es enthält neben der Artbezeichnung sowie der Beschreibung auch die Angaben über Attribute, Verhalten, Beziehungen und Regeln [27]. Ameling [28] unterscheidet zwischen Business-Object-Typen und Business-Object-Instanzen. Ersteres kann mit einer Klasse verglichen werden und beschreibt, welche Elemente sich in dem Business Object befinden können und wie die Struktur aufgebaut ist. Business-Object-Instanzen sind analog zu den Instanzen einer Klasse.

Tabelle 1. Verwendung von Business Objects für betriebliche Anwendungssysteme

Fachgebiet	Teilgebiet	Business Object	Beispiele	Literatur
Informatik	(1) Software-entwicklung	"A business object is a representation of a thing active in the business domain, including at least its business name and definition, attributes, behavior, relationships, rules, policies and constraints." [29]	Mitarbeiter, Produkt, Rechnung, Auftrag, Adresse, Konto, LKW	[24–27, 29–31]
Wirtschafts-informatik	(2) Software-einführung (z. B. ARIS)	„Eine Anwendung besteht aus mehreren Business Objects, die von einem für die Anwendung definierten Geschäftsprozess gesteuert wird." [32]	Auftrag, Artikel, Lieferant, Kunde	[32–34]
Wirtschafts-informatik	(3) Software-orchestrierung	„Ein Geschäftsvorfall (Gesamtprozess) benötigt für seine Abwicklung immer Informationen aus mehreren BO." [35]	Kunde, Produkt, Mitarbeiter, DR-Antrag, Genehmigung, Ablehnung	[35–37]

Ein weiterer Bereich, in dem Business Objects verwendet werden, ist die *Wirtschaftsinformatik*. Das ARIS-Konzept (Architektur integrierter Informations-systeme) wurde zu Beginn der 1990er Jahre entwickelt, um Informationssysteme möglichst vollständig beschreiben zu können. ARIS dient beispielsweise dazu, Unternehmen bei der *Softwareeinführung* und beim Customzing zu unterstützen. Nach Scheer [32] gibt es dazu fünf verschiedene Sichten: Funktionssicht, Organisationssicht, Datensicht und Leistungssicht. Zwischen diesen Sichten werden die Beziehungen modelliert, was die Steuerungssicht darstellt. Das ARIS-Konzept sieht Business

Objects als komplexe Datenobjekte, die Beziehungen zu weiteren Business Objects haben und die charakterisiert sind durch Attribute und zugeordneten Methoden [32]. Die Reihenfolge des Aufrufs der Objekte wird mit Hilfe eines Prozessmodells visualisiert.

Das Business-Object-Konzept wird auch bei der *Softwareorchestrierung* genutzt. Für die Abwicklung eines Geschäftsvorfalls (Gesamtprozess) sind Informationen aus mehreren Business Objects notwendig. Der Geschäftsvorfall *Vertragserstellung* benötigt beispielsweise Attribute der Business Objects *Produkt* (z. B. Bezeichnung, Material), *Kunde* (z. B. Kundennr., Name, Adresse) und *Mitarbeiter* (z. B. Name, Personalnr.) [35]. Bei einer Service-orientierten Architektur stellt dann ein Service dem Anwender eine einheitliche Repräsentation von Business Objects zur Verfügung, wodurch eine homogene Sicht auf die Daten und Operationen eines Business Objects ermöglicht wird [36].

3.2 Ergebnis der Interviews

Alle Interviewpartner (N=6) arbeiten in der IT-Abteilung und sind Administratoren oder Endanwender von betriebswirtschaftlichen Anwendungssystemen. Drei der Befragten hatten *keine* Vorstellung, was sich hinter dem Begriff verbirgt. Drei weitere der Befragten verorteten das Business Object in den Bereich der ERP-Systeme. Bei einem dieser drei Befragten wird SAP im Unternehmen eingesetzt und dieser ordnete den Begriff in das Umfeld von SAP ein. Ein weiterer Interviewpartner (kein Einsatz von SAP in diesem Unternehmen) gab eine beispielhafte Erklärung: „Ein Geschäftsobjekt kann (…) sowohl ein Formular-Inhalt sein (zum Beispiel von einem Urlaubsantrag) wie auch ein Dokument – zum Beispiel ein Kundenauftrag oder eine Rechnung.“ Für diesen Befragten ist ein Business Object „(…) eine Instanz, und Objekte gehören zu Objektklassen und das ist dann ein Mitarbeiter.“

Anhand der Antworten wird deutlich, dass Administratoren oder Endanwender entweder keine oder nur eine vage Vorstellung von einem Business Object haben. Die drei Befragten, die überhaupt eine Vorstellung hatten, stellen allerdings keinen Bezug zum Bereich der Software*entwicklung* her, aus dem das Konzept ursprünglich stammt. Lediglich ein Befragter spricht die Kategorisierung von Unternehmensinformationen an und differenziert zwischen Klassen und Instanzen. Zusammenfassend kann festgestellt werden, dass der Begriff des Business Objects in der Praxis von IT-Abteilungen anscheinend keine wahrnehmbare Bedeutung hat. Aufgrund der sehr geringen Aussagekraft der Interviews wurden nach sechs Durchläufen keine weiteren Daten aus dieser Quelle erhoben.

3.3 Ergebnisse der Systemanalyse der SoNBO-Anwendungen

Wie zuvor erwähnt, gibt es bereits zwei SoNBO-Anwendungen, die die Informationen aus verschiedenen betrieblichen Anwendungssystemen integrieren. In beiden Anwendungen werden die Unternehmensinformationen als *Business-Object-Typen* (z. B. Kunde, Lieferant, Angebot, Auftrag, Rechnung) klassifiziert. Dies dient dazu, dass für jeden Typ die Attribute (z. B. Name) und die Beziehungen in der

Administrationsoberfläche konfiguriert werden. Dabei wählt der Administrator das Feld in dem betriebswirtschaftlichen Anwendungssystem aus, das den Wert (z. B. Name) enthält. Wird vom Endanwender ein konkretes Business Object aufgerufen, wird anhand der Konfigurationen von der Anwendung geprüft, welche Attribute (falls vorhanden) angezeigt werden und welche Beziehungen (falls vorhanden) geladen werden. Diese Informationen werden dem User im Frontend angezeigt. Somit wird die Klassifizierung der Unternehmensinformationen in den Business-Object-Typen sowie in den konkreten *Business-Object-Instanzen* transparent gemacht. Die Klassifizierung der unternehmensinternen Informationen in die Typen ist zum einen durch die technischen Gegebenheiten in den betriebswirtschaftlichen Anwendungssystemen entstanden und zum anderen durch die Anforderungen der User.

3.4 Zusammenfassung Status Quo

Zusammenfassend lässt sich festhalten, dass das Business-Object-Konzept für die objektorientierte betriebswirtschaftliche Anwendungsentwicklung entwickelt wurde. Die OMG leitete die Entwicklung dieses Konzeptes in die Wege. Ein Business Object ist ein logisches Abbild eines betriebswirtschaftlichen Sachverhaltes [38]. Dies wurde von ARIS für die Beschreibung von Softwaresystemen aufgegriffen, um z. B. Software in Unternehmen einzuführen. Für die Softwareorchestrierung wird ebenfalls auf das Business-Object-Konzept zurückgegriffen. Gemeinsam ist allen drei Bereichen, dass Business Objects zur Strukturierung und Kategorisierung von Unternehmensinformationen verwendet werden und dadurch Aufbau und Definition ähnlich sind, sie sich allerdings im jeweiligen Zweck unterscheiden. Außerdem geben Beiträge aus dem Bereich der Softwareentwicklung auch als Beispiel für ein Business Object eine Adresse an, was in den anderen beiden Bereichen (Softwareeinführung und Softwareorchestrierung) als Attribut betrachtet wird. Dies verdeutlicht, dass die Softwareentwicklung einen funktionalen Zweck in den Business Objects sieht, während die anderen beiden Bereiche stärker auf die Bedürfnisse des Unternehmens fokussiert sind. Weiterhin ist den drei Bereichen gemeinsam, dass sich die Kategorisierung auf das Backend bezieht (Zielgruppe: Entwickler, Administratoren), diese Strukturierung allerdings dem Endanwender nicht transparent gemacht wird. Der SoNBO-Ansatz für die Integration von Unternehmensinformationen greift die Strukturierungsmöglichkeit auf, benutzt diese für die Entwicklung und Administration der SoNBO-Anwendung und macht diese auch für den Endanwender transparent.

Offen bleibt in der Literatur, ob es sich bei einem Business Object um eine *Klasse* oder eine *Instanz* bzw. um einen Überbegriff der beiden handelt.

4 Social Business Objects

In diesem Abschnitt wird die Entwicklung einer Definition für Social Business Objects beschrieben. Dazu werden zunächst auf Basis von Abschnitt 3.3 Beispiele eines Social Business Objects sowie deren Charakterisierung aufgeführt, um anschließend auf Basis von Abschnitt 3 eine Definition abzuleiten.

Wie eingangs erwähnt, gibt es graphenbasierte Ansätze zur Integration im Bereich Semantic Web, die auf Unternehmen übertragen werden können: Linked Enterprise Data und Enterprise Knowledge Graph [14, 17]. Ein wichtiges Element ist dabei die Ontologie, die die Zusammenhänge der Informationen abstrakt in „Concepts" (Klassen) beschreibt [39], und der „Knowledge Graph", der die Informationen als Instanzen darstellt. Ein Knowledge Graph ist demnach eine Ontologie verbunden mit einer Faktenbasis (z. B. Datenbank). Eine Ontologie [39] beinhaltet eine Taxonomie sowie eine Konzeptualisierung, die den Bereich definiert, den eine Ontologie beschreiben soll.

4.1 Beispiele und Charakterisierung

Im SoNBO-Ansatz ist das Social Business Object die Instanz einer Klasse, die in Anlehnung an ein Concept in einer Ontologie [39] auch Social Concept genannt wird. Die Unternehmensinformationen (Instanzen, Social Business Objects) werden so strukturiert, dass sie in Kategorien (Klassen, Typen, Social Concepts) zusammengefasst werden können. Ein Social Concept (Mitarbeiter) kann auch Social Subconcepts (Sachbearbeiter, Geschäftsführer, Ingenieur) beinhalten.

Tabelle 2. Beispiel und Charakterisierung eines Social Business Objects (SBO)

	Name	Social Concept	Attribut	Beziehung zu	Methode
Beispiel 1	*Peter Müller*	*Mitarbeiter*	*Alter: 31 Position: Sachbearbeiter*	*1725*	*Position ändern; Nachricht über geränderte Position senden*
Beispiel 2	*1725*	*Ausgehende Rechnung*	*Datum: 02.04.17 Betrag: 50.000,00*	*Peter Müller*	*Rechnungs-datum ändern; Folgen*
Eigenschaft	Ein SBO hat einen Namen.	Ein SBO gehört zu einem Social Concept	Ein SBO hat Attribute. Verschie-dene SBO können nach Attributen gefiltert und sortiert werden.	Ein SBO steht in einer Beziehung zu beliebig vielen weiteren SBO.	Ein SBO bietet CRUD und soziale Funktiona-litäten/ Methoden.

CRUD = create, read, update, delete

Ein Social Business Object ist beispielsweise ein Eintrag in einer (relationalen) Datenbank, während das Social (Sub)Concept die Kategorie des zugehörigen Datenbankeintrages ist. Ein Social Business Object kann sich aus mehreren Datenbankeinträgen zusammensetzen. Ein Social Business Object hat einen Namen und gehört zu einem Social Concept (vgl. Tabelle 2). Für das Social Concept wird identifiziert, welche Attribute (z. B. Alter) es gibt. Die Attributwerte (z. B. 31) werden beim Aufrufen eines Social Business Objects durch den User geladen. Der User kann verschiedene Business Objects nach den Attributen filtern. Weiterhin stehen Social Business Objects in einer Beziehung zu weiteren Social Business Objects. Diese Information befindet sich in dem Datensatz des Social Business Objects. Außerdem hat ein Business Object verschiedene Methoden. Zum einen gibt es CRUD-Funktionalitäten. Ein User kann ein Business Object in einer SoNBO-Anwendung erstellen (create). Weiterhin kann er die Informationen lesen (read). Außerdem ist es möglich, ein Social Business Object zu aktualisieren (update) und auch zu löschen (delete). In den beiden existierenden SoNBO-Anwendungen ist *read* von größter Bedeutung. Auch *update* und *create* (Urlaubsantrag erstellen) wurden implementiert. *Delete* wurde noch nicht realisiert, ist aber prinzipiell auch umsetzbar.

Weiterhin werden für die Gestaltung des Frontends soziale Funktionalitäten benutzt, wie sie aus Kollaborationssoftware bekannt sind [40]. Jedes Social Business Object wird mittels eines *Profils* dargestellt. In Kollaborationssoftware ist ein Profil immer eine Person, während in einer SoNBO-Anwendung ein Profil auch eine Rechnung oder ein Angebot sein kann. Auch das *Folgen* ist vorhanden, wobei sich die Social Business Objects nicht selbstständig miteinander verknüpfen müssen, sondern diese Verknüpfung in den Informationen der betrieblichen Anwendungssysteme bereits vorhanden ist und für SoNBO identifiziert und genutzt wird. Außerdem werden Nachrichten von jedem Social Business Object generiert, wenn sich Änderungen ergeben (z. B. Rechnung wurde vom Kunden bezahlt). Diese Nachrichten werden den umliegenden Social Business Objects in deren Neuigkeiten (Activity Stream) angezeigt.

Die Identifikation der Social Concepts und des Netzwerkes erfolgen durch eine initiale Analyse der betrieblichen Anwendungssysteme und den Anforderungen der User. Dabei gibt es keine richtige oder falsche Kategorisierung der Unternehmensinformationen in Social Concepts. Das Ziel besteht darin, dem User eine sinnvolle Strukturierung transparent zu machen und nicht eine für das Backend logische Struktur zu erstellen.

4.2 Definition

Anhand der Beispiele und der Charakterisierung von Social Business Objects ist zu erkennen, dass diese Business Objects sind. Dadurch besitzen sie die Natur und Struktur von Business Objects und werden um soziale Funktionalitäten erweitert. Allerdings unterscheiden sie sich im Zweck. Während das Business-Object-Konzept vor allem zur Kategorisierung der Unternehmensinformationen im Backend verwendet wird, wird das Social-Business Object-Konzept als Erweiterung verwendet, um die Kategorisierung dem Endanwender transparent zu machen und diesem so eine Informationsexploration zu ermöglichen. Grundlegend dafür ist die Graphenstruktur (= Unternehmensnetzwerk), in der die Beziehungen zwischen den (Social) Business Objects beschrieben werden.

Definition. Social Business Objects sind konkrete Instanzen von Social Concepts. Ein Social Concept kann Social Subconcepts beinhalten.
Social Concepts kategorisieren die in den betrieblichen Anwendungssystemen enthaltenen Organisationsinformationen unter Berücksichtigung der technisch vorhandenen Business Objects in den betrieblichen Anwendungssystemen und den Anforderungen einer Organisation. Die Daten eines Social Business Objects sind (zusammengesetzte) Einträge aus Datenbanken.

Alternative Bezeichnungen.
Social Concept ≙ Klasse, Typ, Kategorie
Social Business Object ≙ Instanz
Hinweis: Der Begriff Business Object wird in der Literatur (abweichend von der hier vorgestellten Definition) auch für das abstrakte Konzept der Klasse verwendet. Dies ist nach Ansicht der Autoren nicht zweckdienlich für die Schaffung von Klarheit.

Aufbau. Ein Social Business Object hat einen Namen, Attribute, Beziehungen zu weiteren Social Business Objects und Methoden. Die Methoden beinhalten CRUD und besitzen soziale Funktionalitäten.

Zweck. Die Strukturierung der Unternehmensinformationen durch Social Concepts und Social Business Objects hat das Ziel, diese dem Endanwender transparent zu machen. Die darin enthaltene Information wird für die Gestaltung der Administrationsoberfläche und für die Entwicklung einer SoNBO-Anwendung verwendet.

Da der graphenbasierte SoNBO-Ansatz sich an betriebswirtschaftlichen Herausforderungen orientiert (Schaffung einer umfassenden Sicht auf Unternehmensinformationen), soll im Folgenden der Zusammenhang zu existierenden graphenbasierten Integrationsansätzen hergestellt werden (vgl. Tabelle 3).

Tabelle 3. Vergleich der Terminologie des Semantic Web und SoNBO

Semantic Web	**SoNBO**		
Bezeichnung des Graphen	**Bezeichnung des Graphen**	**Bezeichnung der Knoten**	**Beispiel**
Ontologie	Social Network of Concepts	Social Concept	Ausgehende Rechnung
Knowledge Graph = Ontologie & Faktenbasis	Social Network of Business Objects	Social Business Object	1725

Im SoNBO-Ansatz wird der abstrakte Graph, der den Aufbau der Unternehmensinformationen beinhaltet und was im Semantic Web als Ontologie bezeichnet wird, *Social Network of Concepts* genannt. Ein Knoten wird dabei als *Social Concept* bezeichnet. Das Netzwerk mit den konkreten Instanzen wird Social Network of Business Objects genannt, wobei ein konkreter Knoten als Social Business Object bezeichnet wird.

An dieser Stelle sei erwähnt, dass der Graph im SoNBO-Ansatz streng genommen „Social Network of *Social* Business Objects“ heißen müsste. Aus Gründen der

Einfachheit wird auf das „Social“ vor „Business“ verzichtet, wobei es aber bei der Knotenbezeichnung verwendet wird, um sich von den Business Objects abzugrenzen.

5 Zusammenfassung und weitere Forschung

Der vorliegende Beitrag beschreibt den Status Quo des Verständnisses zum Begriff „Business Object“. Das Konzept kommt aus der Informatik (Softwareentwicklung von betriebswirtschaftlichen Anwendungssystemen) und wurde von der Firma SAP für ihre ERP-Software aufgegriffen und damit bekannt gemacht. Die Wirtschaftsinformatik verwendet das Konzept zudem für Softwareeinführung und Softwareorchestrierung. Der SoNBO-Ansatz zur Integration von Informationen benutzt ebenfalls das Business-Object-Konzept. Gemeinsam ist diesen vier Ansätzen das grundlegende Verständnis eines Business Objects als Mittel zur Strukturierung von Unternehmensinformationen. Allerdings verstehen die ersten drei Ansätze diese Kategorisierung für den *Softwareentwickler* und der SoNBO-Ansatz benutzt das Konzept, um dem *Endanwender* diese Kategorisierung transparent zu machen und somit eine Informationsintegration umzusetzen. Dazu werden die Social Business Objects miteinander zu einem Netzwerk verknüpft. Das Business Object erhält zusätzlich soziale Funktionalitäten, wodurch der Begriff des Social Business Objects entstand. Außerdem differenziert der SoNBO-Ansatz zwischen einem Social Business Object als konkrete Instanz und einem Social Concept als beschreibende Klasse.

Durch die nun erfolgte Begriffsbildung auf der Grundlage einer Literaturrecherche, Interviews und einer Systemanalyse kann die SoNBO-Forschung in der Folge fortgesetzt werden, indem auf der Basis einer trennscharfen Terminologie an einem Framework gearbeitet wird, das eine Übertragbarkeit von SoNBO auf beliebige Unternehmen ermöglicht. Die Herausforderung besteht darin, das Social Network of Concepts eines konkreten Unternehmens zu identifizieren, um die bestehende SoNBO-Anwendung anhand der Anforderungen dieses Unternehmens zu „konfigurieren“.

Somit stellt der Beitrag einen Erkenntnisgewinn in zweierlei Hinsicht dar. Zum einen wurde herausgearbeitet, welches Verständnis in der wissenschaftlichen Literatur zu einem Business Object vorhanden ist (Status Quo). Weiterhin wurde ein Vorschlag für eine fundierte Definition des Begriffs „Social Business Objects“ entwickelt, der eine notwendige Voraussetzung für die weitere Forschung im Bereich SoNBO darstellt.

Literaturverzeichnis

1. Schwarze, J.: Einführung in die Wirtschaftsinformatik. NWB Verlag (2000).
2. Davenport, T.H.: Putting the Enterprise into the Enterprise System. Harv. Bus. Rev. 76, 121–131 (1998).
3. Williams, S.P.: Das 8C-Modell für kollaborative Technologien. In: Schubert, P. and Koch, M. (eds.) Wettbewerbsfaktor Business Software. pp. 11–21. Hanser, München (2011).
4. Schubert, P., Glitsch, J.H.: Adding Structure to Enterprise Collaboration Systems: Identification of Use Cases and Collaboration Scenarios. Procedia Comput. Sci. 64, 161–169 (2015).

5. Schwade, F., Schubert, P.: Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities. In: 50th Hawaii International Conference on System Sciences (HICSS). pp. 401–410. , Hilton Waikoloa Village, Hawaii, USA (2017).
6. Weske, M.: Business Process Management: Concepts, Languages, Architectures. Springer (2012).
7. Schmidt, A., Otto, B., Österle, H.: Unternehmensweite Stammdatenintegration. Wirtschaftsinformatik Manag. 5, 46–52 (2010).
8. Linthicum, D.S.: Next Generation Application Integration. Addison Wesley, Boston, MA (2003).
9. Holten, R.: Integration von Informationssystemen. Wirtschaftsinformatik. 45, 41–52 (2003).
10. Jung, R.: Architekturen zur Datenintegration. Deutscher Universitätsverlag, Wiesbaden (2006).
11. Liebhart, D.: SOA goes real. Serviceorientierte Architekturen erfolgreich planen und einführen. Carl Hanser Verlag, München, Wien (2007).
12. Linthicum, D.S.: Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide. Addison Wesley, Boston, MA (2009).
13. Krcmar, H.: Einführung in das Informationsmanagement. Springer, Berlin, Heidelberg (2015).
14. Auer, S., Petzsch, R., Unbehauen, J.: Datenintegration im Unternehmen mit Linked Enterprise Data. In: Pellegrini, T., Sack, H., and Auer, S. (eds.) Linked Enterprise Data. pp. 85–101. Springer, Berlin, Heidelberg (2014).
15. Gewehr, B., Gebel-Sauer, B., Schubert, P.: Social Network of Business Objects (SoNBO): An Innovative Concept for Information Integration in Enterprise Systems. Procedia Comput. Sci. 121, 904–912 (2017).
16. Götz, F., Gebel-Sauer, B.: Vom CRM-System ins Social Network of Business Objects (SoNBO): Entwicklung eines Prototyps für eine innovative Informationsintegration. In: Multikonferenz Wirtschaftsinformatik (MKWI) 2018. pp. 1995–2006. , Lüneburg (2018).
17. Galkin, M., Auer, S., Scerri, S.: Enterprise Knowledge Graphs: A Backbone of Linked Enterprise Data. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence. pp. 497–502. IEEE Computer Society, Omaha, NE, USA (2016).
18. Walsham, G.: Interpretive case studies in IS research: nature and method. Eur. J. Inf. Syst. 4, 74–81 (1995).
19. Thomas, G.: A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. Qual. Inq. 17, 511–521 (2011).
20. Kuechler, B., Vaishnavi, V.: On theory development in design science research: anatomy of a research project. Eur. J. Inf. Syst. 17, 489–504 (2008).
21. Torraco, R.J.: Writing Integrative Literature Reviews: Guidelines and Examples. Hum. Resour. Dev. Rev. 4, 356–367 (2005).
22. vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A.: Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. Commun. Assoc. Inf. Syst. 37, 205–224 (2015).
23. Williams, S.P., Schubert, P.: Connecting Industry: Building and Sustaining a Practice-based Research Community. In: Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS). pp. 5400–5409. , Hilton Waikoloa Village, Hawaii, USA (2017).
24. Weske, M.: Business-Objekte: Konzepte, Architekturen, Standards. Wirtschaftsinformatik. 41, 4–11 (1999).
25. Seubert, M.: Business-Objekte und objektorientiertes Prozeßdesign. In: Becker, J., Rosemann, M., and Schütte, R. (eds.) Referenzmodellierung: State-of-the-Art und

Entwicklungsperspektiven. pp. 107–128. Springer-Verlag Berlin Heidelberg GmbH, Berlin, Heidelberg (1998).
26. Zimmermann, F.-O.: Betriebliche Informationssysteme in virtuellen Organisationen. Deutscher Universitäts-Verlag, Bamberg (1999).
27. Eicker, S., Nietsch, M.: Standards zum objektorientierten Paradigma. Wirtschaftsinformatik. 41, 358–370 (1999).
28. Ameling, M.: Systemunterstützung für den Abgleich von Geschäftsobjekten zwischen Anwendungsservern über Web Services, (2009).
29. Casanave, C.: Business-Object Architectures and Standards. In: Sutherland, J., Patel, D., Casanave, C., Hollowell, G., and Miller, J. (eds.) Business object Design and Implementation. pp. 7–28. Springer-Verlag, Berlin Heidelberg New York (1995).
30. Frank, U.: Component Ware: Software-technische Konzepte und Perspektiven für die Gestaltung betrieblicher Informationssysteme. Inf. Manag. Consult. 14, 11–18 (1999).
31. Sims, O.: The OMG Business Object Facility and the OMG Business Object. In: Sutherland, J., Patel, D., Casanave, C., Hollowell, G., and Miller, J. (eds.) Business object Design and Implementation. pp. 37–46. Springer-Verlag, Berlin Heidelberg New York (1995).
32. Scheer, A.-W.: ARIS-Modellierungsmethoden, Metamodelle, Anwendungen. Springer, Berlin; Heidelberg (1998).
33. Nüttgens, M., Zimmermann, V.: Geschäftsprozeßmodellierung mit der objektorientierten Ereignisgesteuerten Prozeßkette (oEPK). In: Maicher, M. and Scheruhn, H.-J. (eds.) Informationsmodellierung. Harzer wirtschaftswissenschaftliche Schriften. pp. 23–35. Deutscher Universitätsverlag, Wiesbaden, Wiesbaden (1998).
34. Scheer, A.-W., Nüttgens, M., Zimmermann, V.: Objektorientierte Ergenisgesteuerte Prozeßkette (oEPK) - Methode und Anwendung -. , Saarbrücken (1997).
35. Bleiholder, J., Schmid, J.: Definition von Datenarten zur konsistenten Kommunikation im Unternehmen. In: Hildebrand, K., Gebauer, M., Hinrichs, H., and Mielke, M. (eds.) Daten- und Informationsqualität. pp. 143–156. Vieweg+Teubner Verlag, Wiesbaden (2008).
36. Dietzsch, A., Goetz, T.: Nutzen-orientiertes Management einer Service-orientierten Unternehmensarchitektur. In: 7. Internationale Tagung Wirtschaftsinformatik (WI). pp. 1519–1538 (2005).
37. Schmidt, W., Fleischmann, A., Gilbert, O.T.: Subjektorientiertes Geschäftsprozessmanagement. HMD Prax. der Wirtschaftsinformatik. 46, 52–62 (2009).
38. Schmidt-Volkmar, P.: Betriebswirtschaftliche Analyse auf operationalen Daten. GWV Fachverlage GmbH, Wiesbaden (2008).
39. Guarino, N., Oberle, D., Staab, S.: What Is an Ontology? In: Staab, S. and Studer, R. (eds.) Handbooks on Ontologies. pp. 1–20. Springer, Dordrecht, Heidelberg, London, New York (2009).
40. Schubert, P.: Joint Work and Information Sharing in the Modern Digital Workplace: How the introduction of “social” features shaped Enterprise Collaboration Systems. In: Collaboration in the Digital Age: How Technology Enables Individuals, Teams and Businesses. pp. 45–59. Springer Verlag, Berlin, Heidelberg (2019).

Designing a Reference Model for Digital Product Configurators

Sarah Hönigsberg[1], Christoph Kollwitz[1], and Barbara Dinter[1]

[1] Chemnitz University of Technology, Business Information Systems I, Chemnitz, Germany
{sarah.hoenigsberg, christoph.kollwitz, barbara.dinter}@wirtschaft.tu-chemnitz.de

Abstract. Since the manufacturing industry is nowadays facing increasingly heterogeneous customer requirements, digital product configurators (DPC) have become a popular means for integrating customers into organizational value creation. DPC are information systems, which serve as a frontend to the customer and enable the individualization of products. The design of such a DPC is time consuming, expensive and lacks appropriate models offering guidance for its development. The paper at hand addresses these issues by providing a reference model (RM) for DPC development. The model has been constructed by means of an extensive literature review and was subsequently demonstrated and evaluated in a real world scenario. In order to ensure a flexible and individual development of company specific DPC the RM includes adaptation mechanisms. Therefore, our research provides a first building block to the endeavor of facilitating or even automating DPC development.

Keywords: Reference Model, Product Configurator, Information System Modeling, Software Engineering.

1 Introduction

Nowadays organizations experience a business environment, where the focus has shifted from mass produced goods to individual products tailored to customer needs [1]. This trend towards personalization was mainly driven by the development of information and production technology [2]. The customer's active involvement in the product design serves as an example, whereby so-called digital product configurators (DPC) enable user-friendly and fast communication and the collection of customer preferences regarding the product customization [3]. Well-known examples are car configurators enabling customers to personalize their car by choosing e.g. interior, engine types, or color. Although DPC promise many advantages, such as reducing process cycle times through automation, decreasing faulty orders, or increasing sales [4], they exhibit downsides, too. High development efforts and costs as well as the difficult maintainability of DPC have to be emphasized [5, 6]. Especially for small organizations, developing and deploying a DPC is challenging and time-consuming. At the same time the DPC has a huge impact on their business and becomes a key system that centralizes the exchange of product, tender, and customer data [7]. This effect

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

occurs because the DPC supports one of the company's core processes (selling the product in combination with customer contact) and thereby requires access to adjacent systems such as customer relationship management, enterprise resource planning, or product data management systems [8]. In small companies without such information systems (IS), parts of these IS must be integrated into the DPC, e.g. to store the product specification necessary for the configuration task. This leads to very complex DPC and highly individual company requirements towards it. Missing guidelines for a sound development of DPC constitute a major challenge often resulting in an ad hoc development of these complex IS without conceptual planning and in problems such as the entanglement of software layers, e.g. the graphical user interface (GUI) in the frontend and the business logic in the backend [6].

While researchers have addressed many different aspects of DPC development in an isolated way such as product knowledge modeling, principles for designing GUI, or reasoning algorithms [9] and studies have been conducted examining several hundred DPC [10], a consolidation of the results from a variety of research fields such as economics, innovation or production research is missing. Therefore, a transformation of the acquired knowledge from the different research fields to the IS research as a holistic conceptual view for the development of DPC is absent. On the one hand, the motivation to build a guidance supporting the DPC software developers lies in the importance of DPC in the modern business environment and on the other hand in the challenges of their development. To guarantee that this guidance is suitable for as many organizations as possible, an artifact representing the best practices for DPC is needed. This artifact should be reusable and universally applicable for many cases. Reference models (RM) are particularly suitable for such purposes. They support system design and development since they streamline the design process and facilitate the implementation by providing best practices, recommendations and represent the knowledge in a particular field [11].

Taking the mature research field of DPC into account, the paper at hand aims to consolidate prior research findings from the last decades to this novel, overarching artifact. The authors develop a RM for DPC as an IS on a conceptual level, based on a secondary literature study including prior literature reviews and studies. This approach leads to the following research question (RQ): *How does a DPC RM need to be built in terms of design decisions and scope to make existing knowledge accessible to IS scholars and practitioners?* To answer this RQ, the remainder of the paper is organized as follows: The (1) introduction with the problem definition is followed by the (2) DPC foundations, and then the (3) research approach is described. Afterwards, the (4.1) DPC characteristics are derived from literature. Using this foundation the RM is (4.2) built, (5) applied, and evaluated. The paper ends with the (6) discussion of the findings and an (7) concluding section.

2 Foundations

Product configuration as a sub-class of configuration in general, constitutes a particular type of design activity. The product to be designed is composed of a set of predefined components that can only be combined in a certain set of combinations, which are called solution space [12]. In order to address and facilitate this often complex task, IS are used, which contain the knowledge about the components and the rules how these can be combined [13]. DPC can support quotation, marketing, and rapid manufacturing of customized products in various industries from car manufacturers to software vendors [9, 13]. The application of DPC in this context has already been discussed in academia and practice since the late 1970s [14]. In DPC research, topics ranging from system design to the impact of configuration have been discussed [9]. Research on this topic, however, often revolved around user or customer interaction. Aspects such as complexity perception, customer satisfaction, and user interaction patterns were addressed [5] in order to optimize the experience with DPC. Not only the experience or fun of configuring products was discussed, but also how DPC as an advisory systems can help customers find the right product to meet their needs [8]. In contrast, a study in which 500 online business-to-customer DPC were evaluated and categorized by features, indicates that they are often still parameter-based, decision-focused, and technology-oriented [10]. This leads to the conclusion that the customer experience is widely discussed in theory, however not properly addressed in the real world. In addition, good and bad practices of development were presented, e.g. a study of 111 web sales DPC raised questions about the presentation of configuration options, the handling of restrictions, and the support of configuration processes [6].

A more comprehensive view of DPC is provided in Subsection 4.1, where a systematic literature review is carried out to analyze the domain as an integral part of reference modeling. The reference modeling itself is described in Section 3 below.

3 Research Approach

To design the RM for DPC, the development process according to [15] has been used in conjunction with an evaluation framework for RM [16] and the design science research approach [17]. In contrast to a (project or enterprise) specific model that would describe a specific DPC, a RM for DPC describes this class of IS in an idealized way [15]. The method of reference modeling can be divided into two phases: (A) construction of the RM and (B) usage of the resulting RM [15]. Both phases are each divided into four individual steps. Common construction steps are (A.1) problem definition, (A.2) conceptualization, (A.3) evaluation, and (A.4) maintenance of a RM. Typical usage steps include (B.1) selection, (B.2) customization, (B.3) integration, and (B.4) application of the RM to solve a problem [15]. The paper at hand addresses the steps (A.1) to (A.3) and (B.2). For (A.1) a systematic literature review was conducted [18, 19] (cf. Subsection 4.1). We have focused on existing research results and DPC studies to consolidate the previous extensive research into a secondary study [20]. After the DPC characteristics have been derived, the (A.2) conceptualization follows, in which adequate views, such as a functional view and the corresponding modeling language, are defined (cf. Subsection 4.2). The multiperspectivity constitutes an important modeling concept, meaning that the same IS is described from multiple viewpoints. In addition to the views, the adaptation mechanisms are developed in the conceptualization step. In general, adaptive RM are models that can be adjusted to a particular application and/or condition. Adaptation mechanisms are needed to guarantee that the RM is suitable for different situations and to provide a solution to the reference modeling dilemma, which states: the more specific a RM addresses an organizational problem the more likely the company will choose it, but the more specific a RM is the fewer companies are addressed by it [21]. The focus on adaptive reference modeling in the paper at hand is justified by the wide range of DPC that need to be accommodated (cf. Subsection 4.1). There are two kinds of adaptive RM: the configurative and the generic adaptive RM [21]. Configurative models contain rules on how the model is to be adjusted to a specific situation via configuration parameters, hereby they are mostly easier to use but harder to construct. Generic adaptable models provide a framework but can only be adjusted manually and thereby are easier to construct, however require more effort in use [21]. Since the goal of the RM for DPC is to reduce development efforts and to imprint obtained knowledge into the RM, we have opted for a configurative RM. To guarantee the best possible fit of the generated specific model we have additionally included generic adaptation mechanisms and thereby keep the possibility of altering the specific model after the model configuration.

After the construction, the RM has been demonstrated by (B.2) customizing it for a real world scenario (cf. Subsection 5.1). The final step of reference modeling presented in this paper encompasses the (A.3) evaluation (cf. Subsection 5.2). On the one hand, for the analytical part we have used the Guidelines of Modeling (GoM) [22] and on the other hand, for the empirical part, we have conducted expert interviews. In the second round of the empirical evaluation, the RM has been used in an organizational setting in the ongoing research project PROFUND to build and deploy a cross-organizational DPC for companies without pre-existing DPC.

4 Construction of the RM

In this section, the paper at hand addresses the first two steps of the construction phase of reference modeling. The following literature review defines the problem and analyzes the domain of the RM. By deriving the DPC characteristics, it builds the foundation for the conceptualization in Subsection 4.2. During the conceptualization, the views and adaptation mechanisms are built.

4.1 Deriving the DPC Characteristics from Literature

Although some reviews and studies partly focus on DPC as an IS from a design and development perspective, to our best knowledge the holistic conceptual view for the development of DPC is neglected. Our secondary literature study based on [18, 19] aims to fill this gap. It includes prior literature reviews and studies describing an extensive amount of DPC in order to gain a comprehensive overview (cf. Table 1 for a fact sheet) and addresses the questions of how the DPC is described as an IS, which perspectives exist, and which overall picture emerges. Through the analysis of the papers we have developed ten different concepts in several rounds by axial coding, in order to summarize and aggregate the DPC characteristics. The morphological box is used in accordance with the general morphological analysis [23] to break down the complex properties of DPC into individual concepts and aspects. This creates a clear structure and a solution space that is easy to interpret. The disadvantage of the morphological box for this review - that common or valid combinations of aspects are not recognizable - is disregarded in favor of concept centricity and easy readability of the results. The percentages for each aspect in the morphological box indicate in how many documents of the literature corpus the specific aspect has been described as shown in Table 1.

Table 1. Morphological Box with DPC Characteristics

Concepts	Aspects					
(1) Purpose of configurator	Product development (25,93%)	Product configuration & sales (92,59%)	Customer preference formation (25,93)	Data collection & data analysis (33,33%)		
(2) Product type	Tangible (74,07%)	Intangible (40,74%)				
(3) Addressee of configurator	External (81,48%)	Internal (33,33%)				
(4) Production concept	Development- / Engineer-to-Order (29,63%)	Make-to-Order & Assemble-to-Order (55,56%)	Match-to-Order / Locate-to-Order (7,41%)			
(5) Type of configuration	Parameter-based & component-based (96,30%)	Requirement-based (55,56%)	CAD (25,93%)			
(6) Standard functions (GUI)	Active input (92,59%)	Product information (77,78%)	Visualization (70,37%)	Videos & animations (33,33%)	Sound (3,70%)	
	Price calculation & display (51,85%)	Guided process & navigation (51,85%)	Save & share (29,63%)	Feedback (62,96%)		
(7) Support functions	Validation (70,37%)	Recommendation (59,26%)	Basic configuration (51,85%)	Comparison (22,22%)		
	Automated solution generation (44,44%)	Simulation (11,11%)	Prototype & 3D print (11,11%)	Community & expert feedback (40,74%)		
(8) Knowledge repres. / logic	Rule-based-reasoning (33,33%)	Case-based-reasoning (25,93%)	Model-based-reasoning (37,04%)			
(9) Data models / management	Customer / session data (44,44%)	Configuration rules (62,96%)	Product model (74,07%)	Configuration (81,48%)	User interface data (22,22%)	Administration data (14,81%)
(10) System design	Structure (51,85%)	Function (59,26%)	Data / function integration (59,26%)			

Fact sheet - Search Terms: Produktkonfigurator, Konfigurator, Produktkonfiguration, Sales Configurator, Configurator AND Product, Configurator AND Mass Customization, Toolkit AND Mass Customization, Frontend AND Mass Customization, Customer-Company Interaction System. Databases and Search Engines: IEEE Xplore, AIS eLibrary, Wirtschaftsinformatik-Genealogie-Datenbank (wige.net), Web of Science and Google Scholar; searched in titels only. Number of Articles found: 101 initially, 76 relevant after titel-check; 136 additional paper after forward-backward search; total 214, 68 relevant after abstract-check; 27 relevant after full paper check. Definition of relevanc: the DPC is discussed as a IS from a software engeneering / design perspective.

Existing literature addresses various (1) purposes of DPC ranging from product development over sales and customer preference formation to data collection and analysis [8, 24]. The most discussed purpose is product configuration with the intention of selling the product afterwards. The (2) product type a DPC is used for can be tangible, like cars or intangible, like software [8]. The (3) addressee of a DPC ranges from engineer or salesman to customer, including companies as a customer or end customers. To simplify the fact of user group variation, the different users were divided into two groups: external, which means customers or the public, and internal representing employees. DPC for tangible products and those which are used externally are examined most frequently, respectively. The focus of customer integration by DPC lies on Made-to-Order (MTO) and Assemble-to-Order (ATO), but the entire spectrum from Engineer-to-Order (ETO) to Match-to-Order (MCHTO) or Locate-to-Order

(LTO) [25] is represented [26–28], as shown in the (4) production concept. The different production concepts relate to the grade of the customizability of the product ranging from pure customization (ETO) to mass production (MCHTO/LTO). The (5) type of configuration describes, how the product can be altered, e.g. by changing parameters or components or by stating a requirement which presets the parameters and components needed to satisfy this requirement [29]. As our literature corpus indicates, the configuration via parameters and components is by far the most common way to realize product customization with DPC. Many (6) standard and (7) support functions of a DPC, e.g. updating the product visualization or automatic validation [30] could be identified. Some functions such as active input, which means the act of entering, e.g. a configuration parameter, are almost always mentioned but others such as (listening to product) sounds are rarely mentioned. To realize the different customization approaches and to guide the user through the configuration task, DPC need to include the configuration rules and algorithms, which can be multi-tiered and complex [29, 31]. This is addressed by (8) knowledge representation and reasoning for DPC, which typically distinguishes the three categories: rule-based, case-based, and model based. Each approach represents the knowledge in different ways, either by if-then-rules or by case similarity or by using constraints [8]. In addition, various (9) data models and (10) system designs have been discussed in literature, e.g. the distinction of user frontends and technician frontends with their respective corresponding logic [32]. It is notable that researchers discuss the system design concepts less distinctively, which forms a strong contrast in regards to the intensively discussed distinct functions of DPC.

All in all, DPC is an umbrella term encompassing wildly different IS, which only have in common that they are used for product configuration tasks in a broad variety of domains. Due to this, the (cost-) efficient development of DPC remains a challenge, as there is no standard procedure to guide such development. No single perfect DPC design is available because of the variety concerning: the industries they are used in, the products they can be used for, the customer integration point they can be used at, the customization approach they can realize, and even the user they can address. This leads to the conclusion that the RM needs to be specific enough for software developers to guide the development of a specific DPC but at the same time generic enough to be applicable for ETO, MTO, or MCHTO settings in different industries. Therefore, a preferable solution for the aforementioned problem is a configurative RM addressing the variety in this class of IS [cf. 21].

4.2 Building the Views

The Approach. Based on the findings from the literature review the requirements for the RM for DPC were derived and led to some major design decisions. The views of the RM could be determined, since from a development perspective on the results, there is a clear emphasis on DPC functions, architecture, and data models with a strong focus on functionality. Accordingly, the RM needs at least two different views to represent the different DPC aspects: a functional view describing user interaction with the DPC and an architectural view describing the system architecture including the data models in the backend. The literature review indicates that the RM must be adaptable in order to cover the broad field of DPC applications. Figure 1 illustrates how the morphological box was used to build the foundation of the RM. There are three different sections the concepts can be divided into. The first five concepts build the setting options for the modeler represented by a choice board. The standard and support functions can be converted to a functional view and the last three concepts can be converted to an architectural view. The resulting three artifact types are interconnected via RM configuration rules. The components of the artifact are described in detail below.

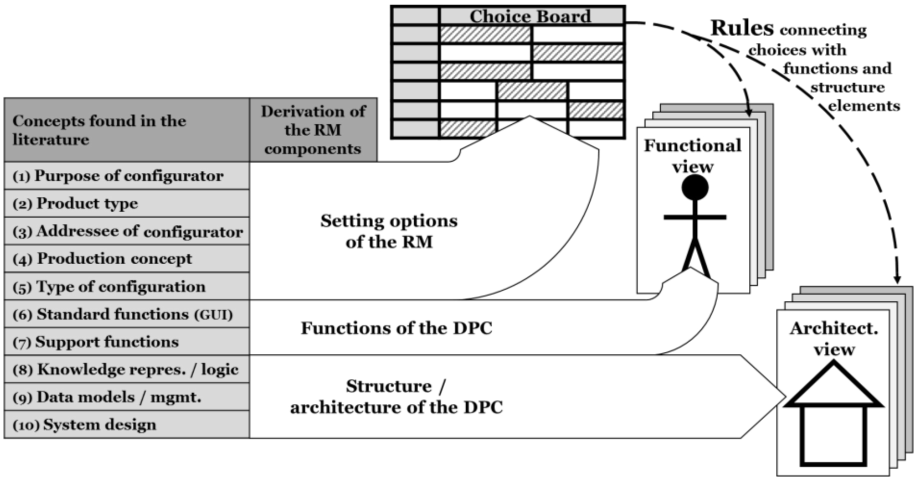


Figure 1. Derivation of the RM Components

Choice Board. What is here called a choice board, is again a morphological box. It is well suited to show configuration dimensions (concepts) of the RM as well as selection options (aspects) in the dimensions. This is a common representation of the multidimensional solution space that illustrates the multitude of variations of RM [33]. There are six questions asked representing the dimensions on the left side of the choice board such as: “Which purpose will the configurator have?”. On the other side, there are predefined answers as selection options like: “Product development” or “Product configuration & sales” (cf. Table 2 in Subsection 5.1). Two kinds of rules were defined in the RM construction. The first one describes the effects of the choice board selections on the two views (cf. rules below) and the second one describes valid combinations in the choice board. It is possible to configure over 270 different versions of the model. An example of a rule is: IF product development THEN NOT Match-to-Order AS production concept. It is suggested to use the choice board by answering the questions in a top-down order, which makes it easier to use.

Functional View and Architectural View. Once the views were identified, we chose adequate modeling languages. In our context, we apply the unified modeling language (UML) Use Case diagram to build the functional view due to two reasons:

First, UML is commonly used in this specific context (describing DPC) and therefore already used in this domain and second, UML is widely accepted and a well-known standard in the field of software engineering [31, 34]. The application of an established modeling standard leads to RM, which are easier to use and avoids starting barriers. Applying a use case diagram has multiple advantages. First, it is possible to build hierarchies of DPC functions like "View product information" as a generalization of "View pictures", "Read description", and so on. Second, the actors included in this diagram type are representing the different DPC users showing the typical interaction with the DPC. During the design process the frequency of mentioning the function in the literature review was taken into account, resulting in mandatory DPC functions in the RM. The full version of the diagram is shown in Figure 2 in Section 5.

For the architectural view an architecture diagram also based on UML was chosen, taking the principle of systematic design into account. This modeling principle demands the inter-model consistency between different views [22]. The condition is fulfilled by using a specified modeling standard such as UML for both views. The architectural view is based on the common three-layer architecture with GUI, application layer, and persistence layer. The view contains components, data elements, and their relations corresponding to the functional view. Both views are linked by a set of rules, so that a selection in the choice board results in changes in both views.

Adaptation Mechanisms and Rules. After determining the representation (modeling language) and the model elements for the views (functions and architectural components), the adaptation mechanisms for the RM were defined. By analyzing the literature, it became clear that dependencies between the DPC functions as well as the architecture and the application field of DPC exist. The RM constitutes a logical representation of these dependencies and therefore becomes a configurator for the DPC as an IS on a conceptual level. The rules for the RM were fully elaborated, resulting in over 100 rules for the configuration and about 40 rules for the generic adaptation. The software developer configures the model by using the choice board. In the first step (the configuration) both views are generated and in the second step (the generic adaptation) individual components are modified if necessary. The result is the recommended functional and architectural scope for a specific DPC. An example of two configuration rules combined is as follows: IF product configuration & sales AS purpose of DPC THEN calculate price information AS function AND price calculation service AS architectural component. Several rules have been defined for each selection option in the choice board. That is, when a selection is made in the choice board, a corresponding set of functions and components is added to the diagrams. In the case that the software developer does not agree with these recommendations, he/she can use the configured model as a starting point and delete or add elements in the model using the generic adaptation rules. Removing, e.g. the price function triggers the generic adaptation rule: IF REMOVE calculate price information AS function THEN REMOVE price calculation service AS architectural component, vice versa. Adding functions and components is limited to the maximum scope of the RM, which is shown in the following section. The demonstration will point out how this RM can be used for a real world scenario.

5 Usage and Evaluation of the RM

In this section, we address the second step of the usage phase and the third step of the construction phase of reference modeling. To perform the customization step of reference modeling, both adaptation variants including the configuration and generic adaptation mechanisms were demonstrated in Subsection 5.1. After this demonstration of the usage, the analytical evaluation to assess the quality and expressiveness of the RM diagrams and the empirical evaluation to assess the suitability and applicability of the RM in real world cases follows in Subsection 5.2.

5.1 Using the RM for DPC

The procedure to use this DPC RM is as follows: First, the DPC developer gathers the initial requirements, second, the DPC developer fills in the choice board with these requirements, third, the RM generates a recommended solution diagram, fourth, the DPC developer uses the generic adaptation to get a better fit of the solution, and finally fifth, the DPC developer uses the specific model to develop a DPC.

For the demonstration and the subsequent evaluation a software development company specialized in DPC development was involved. They described a real case of a prior DPC they had developed for an Austrian manufacturer of customizable front doors. The manufacturer wanted to be able to support its own sales and trading partners in the preparation of quotations. The DPC development company used our choice board to specify the requirements of the manufacturer. To avoid any bias, the developers did not know the rules and logic behind the RM. In Table 2, the selected configuration is shown in the choice board.

Table 2. Choice Board

Dimension / Question	Answer		
Which purpose will the configurator have?	Product development		**Product configuration & sales**
Who will use the configurator?	**External (customer)**		Internal (employee)
Which are additional pursued objectives?	**Customer preference formation**		Data collection & data analysis
Which production concept is pursued?	ETO	**MTO / ATO**	MCHTO / LTO
What type of configuration is needed?	**Parameter / component**	Requirement	CAD
How will the system be integrated?	**Stand-alone**	Data exchange	System integration

The use case diagram shown in Figure 2 includes the maximum scope of functions in the RM, even though not all functions are included in the solution for the manufacturer. This is done for illustration purposes, the legend explains which functions are included and why.

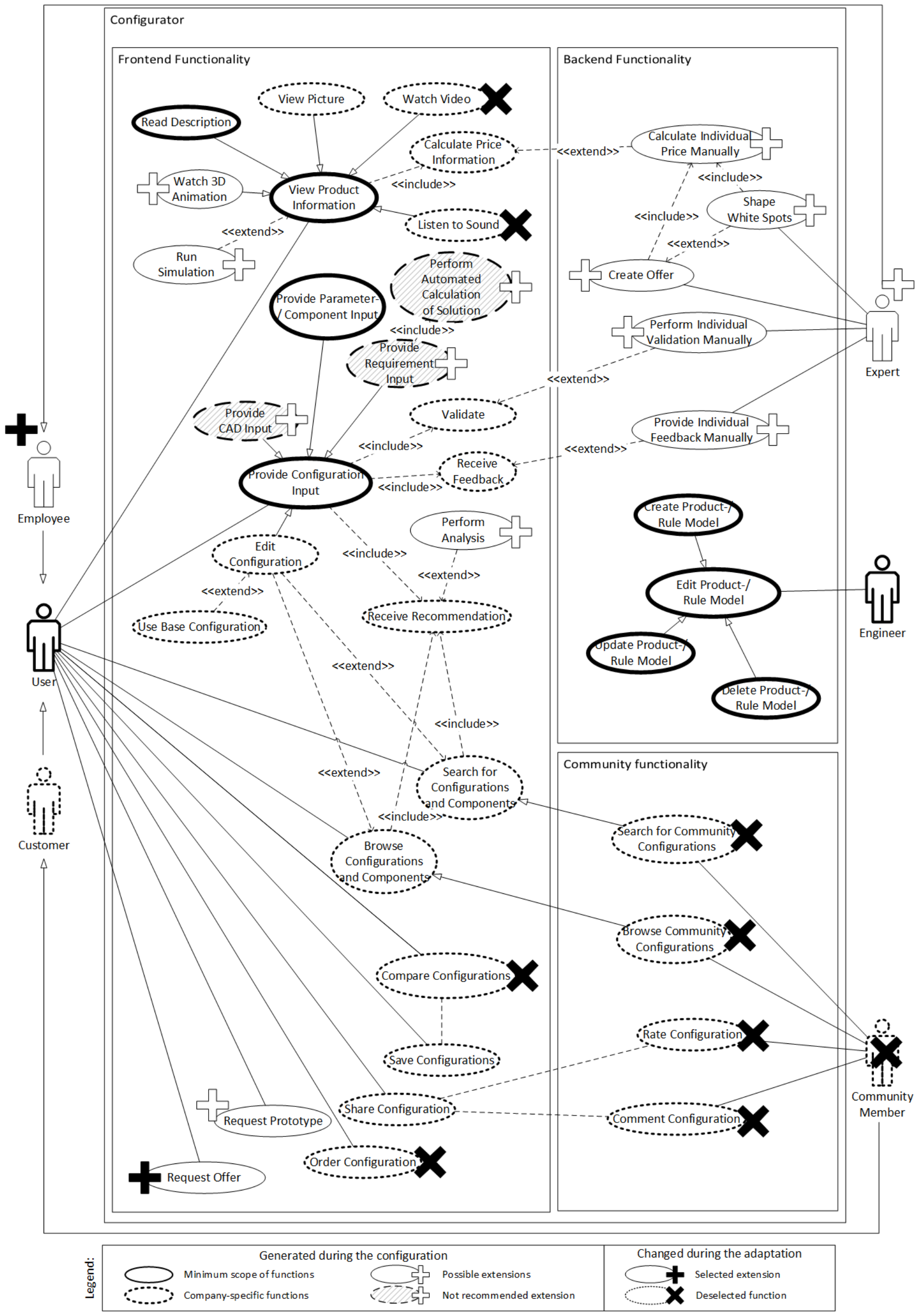


Figure 2. Demonstration of the Functional View

The four symbols on the left side of the legend describe the highlighting of the function explaining the recommendations of the RM after the configuration. The two symbols on the right illustrate the highlighting of the functions, which are added or removed during the general adaptation.

After the model configuration we presented the result to the DPC development company. The developer then had the chance to perform the generic adaptation step to have a closer match to the real DPC of the manufacturer. It is recognizable in Figure 2 that, besides a few other functions, those concerning community integration have been removed whereas the function “request offer” and the employee-actor have been added by the DPC development company. As mentioned before, the architectural view is interconnected with the functional view. Its components correspond with the functions. However, this diagram is not included in this section. Nevertheless, the whole demonstration including both views was performed with the DPC development company and was used for the evaluation afterwards.

5.2 Evaluating the RM for DPC

Analytical Evaluation. For the analytical evaluation, the GoM were used to assess the quality and expressiveness of the RM diagrams. The GoM consist of six principles: (1) construction adequacy, (2) language adequacy, (3) economic efficiency, (4) clarity, (5) systematic design, and (6) comparability [22]. The conclusions of the assessment were the following. (1) and (3) are completely fulfilled, since: first, the problem definition has proven that a RM is a suitable way to address the situation, second, the two views (functional and architectural) are a well-accepted way to describe IS, third, the RM does not include unnecessary content, and fourth, it is considered as economically efficient to build a RM to address a cross-industry problem. The assessment of the latter is based on the fact that in the empirical evaluation the development company estimated an acceleration of DPC development using the RM, which could also be confirmed in the research project PROFUND. Also (4) and (6) were classified as fully satisfied, since UML and UML-related modeling languages were used. However (2) and (5) were rated as partially satisfied, due to the semiformal language and the slightly altered diagram types, although a meta model integration for both views was carried out (though not included in this paper) to guarantee a systematic design.

Empirical Evaluation. At first, the usage perspective was evaluated with the question how suitable the choice board is and whether it is sufficient for the elicitation of requirements in the first stage of DPC development. This was realized by conducting an expert interview with a developer and a requirements engineer in the DPC development company mentioned above. The experts have considered the approach as promising and suitable, but a refinement of the questions and answers is desirable in order to obtain a higher degree of detail in the initial requirements engineering phase. They also identified three major shortcomings of the choice board: First, the external use of the DPC is not divided into the categories of business-to-business and business-to-customer, which can have implications for the frontend design or user handling. Second, the dimension / question for the addressed customization approaches is missing (fit, style, etc.), and third, the rules in the choice board were too strict. While the ruleloosening was realized during development process iterations, the other issues need to be addressed by a broader approach in the RM including, e.g. design suggestions for the frontend or the user handling, e.g. a user login in case of a business-to-business DPC.

At second, the result after using the RM in a real world scenario was assessed. Since the RM was applied to configure a specific model for an already implemented DPC, the resulting model could be compared with the functionality and the architecture of the real system backend of the Austrian front door manufacturer. The comparison was carried out by one of the developers and not by us. The feedback exhibited that a function in the functional view and respectively a component in the architectural view concerning "requesting offers" were missing. As it can be seen in Figure 2, the missing parts were added in the prior RM refinements and are already included in the paper at hand.

At third, the RM was used in the research project PROFUND. The RM was used to accelerate the requirements engineering and the conceptual design of the information technology landscape in the project including a DPC. This led to a prompt conceptual model of the DPC, which was generated by the RM and was used as a foundation for discussion. However, the usage has also shown that it is not facilitating the solution space development and knowledge representation of DPC so far. The usage of the RM in this context will lead to a broader evaluation concentrating on distributed configuration in value chains in an ETO environment and including the whole process of software development. Feedback rounds with to the suggested functionality of the DPC showed that searching for parts, configuring as well as saving the configuration related to the customer data were seen as the most important functionalities, whereby generating offers or sending messages to the customer were deemed irrelevant by the textile companies. The first prototype of the DPC based on the RM has been developed and first feedback indicates that the prototype constitutes a promising approach.

6 Discussion

This paper presents a RM for DPC, which consolidates prior research findings and integrates them into the views, the choice board, as well as the rules and thus forms a novel overarching artifact. The recommendations included in the RM are derived from a comprehensive sample of individual cases and therefore offer general validity. We addressed the missing holistic approach by consolidating prior research viewpoints on DPC such as functionality, architecture, usage, and point of customer integration. While studies like [10] examine a specific type of DPC in a comprehensive way, this paper specifically targets the wide range of DPC to identify the underlying structure and mechanisms of this IS type. Thereby, the presented approach of reference modeling is an example on how to deduce specific features of a RM via a systematic literature review in a mature research field, instead of abstracting from specific models in an inductive way [35].

The evaluation results indicate that the RM is able to accommodate real world requirements successfully and that the RM is applicable across different industries, such as the discrete manufacturing industry (e.g. front doors) as well as the process industry (e.g. technical textiles). The application of the RM in a software development project for small and medium sized enterprises in the textile industry has confirmed the potential benefits of the RM, but has also revealed its limitations. The current RM only allows manual deriving of a specific model. Instead, the model could be automatically generated by a software tool with the option to analyze the usage of the RM for subsequent optimization. In addition, the current RM provides abstract recommendations and no specific support for the actual design and implementation.

Further development could break the RM down to more specific models in order to support the development efforts not only in the initial stages, but also in the subsequent development stages. This purpose could be achieved by, e.g. refining the modules with class diagrams.

7 Conclusion

The paper at hand aims to support organizations in the development of DPC. In a first step we have deduced the characteristics and functions of DPC by conducting a systematic literature review. Considering the wide variety of DPC applications, we have made the design decision that the RM must be adaptive to ensure universality but also specificity. As shown in the demonstration, the adaptive RM constitutes a tool, which can support the development of DPC starting with the requirements engineering stage. The RQ on how to build a DPC RM in terms of design decisions and scope to encompass existing knowledge was answered successfully. The resulting RM is designed to address the problem of high expenses related to the development of DPC, since recommendations for the functional scope and the architecture of the DPC are provided at a very early stage in the development process.

In summary, our research contributes to the knowledge base by applying reference modeling to the field of DPC by making knowledge from research and practice available and thereby usable for developers. Thus, it provides a first building block to the endeavor of facilitating DPC development.

Acknowledgement: *The research described in this paper was supported by a grant from the German Ministry for Research and Education (BMBF), project name: PROFUND grant number 03ZZ0618C.*

References

1. Pine, J., Victor, B., Boynton, A.C.: Making Mass Customization Work. Harv. Bus. Rev. 71, 108–116 (1993)
2. Fogliatto, F., da Silveira, G., Borenstein, D.: The mass customization decade: An updated review of the literature. Int. J. Prod. Econ. 138, 14–25 (2012)
3. Gandhi, A., Magar, C., Roberts, R.: How technology can drive the next wave of mass customization. (2013)

4. Zhang, L.L., Helo, P.T.: Empirical investigation on implications of configurator applications for mass customization. In: Modrak, V. (ed.) Mass Customized Manufacturing: Theoretical Concepts and Practical Approaches. pp. 217–233. Taylor & Francis, London (2016)
5. Franke, N., Piller, F.: Key research issues in user interaction with user toolkits in a mass customisation system. Int. J. Technol. Manag. 26, 578–599 (2003)
6. Abbasi, E.K., Hubaux, A., Acher, M., Boucher, Q., Heymans, P.: The Anatomy of a Sales Configurator: An Empirical Study of 111 Cases. In: Proceedings of the Advanced Information Systems Engineering Conference (CAiSE'2013). pp. 162–177. , Valencia (2013)
7. Forza, C., Salvador, F.: Product configuration and inter-firm co-ordination: An innovative solution from a small manufacturing enterprise. Comput. Ind. 49, 37–46 (2002)
8. Blecker, T., Abdelkafi, N., Kreutler, G., Friedrich, G.: Product Configuration Systems: State of the Art, Conceptualization and Extensions. In: Proceedings of the Maghrebian Conference on Software Engineering and Artificial Intelligence (MCSEAI'2004). pp. 25–36. , Tunis (2004)
9. Zhang, L.L.: Product configuration: a review of the state-of-the-art and future research. Int. J. Prod. Res. 52, 6381–6398 (2014)
10. Walcher, D., Piller, F.: The Customization 500: An International Benchmark Study on Mass Customization and Personalization in Consumer E-commerce. Lulu Marketplace (2011)
11. Rosemann, M., van der Aalst, W.M.P.: A configurable reference modelling language. Inf. Syst. 32, 1–23 (2007)
12. Mittal, S., Frayman, F.: Towards a Generic Model of Configuraton Tasks. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'1989). pp. 1395–1401. , Detroit (1989)
13. Helo, P.T.: Product configuration analysis with design structure matrix. Ind. Manag. Data Syst. 106, 997–1011 (2006)
14. McDermott, J.: R1: A rule-based configurer of computer systems. Artif. Intell. 19, 39–88 (1982)
15. Fettke, P., Loos, P.: Referenzmodellierungsforschung. Wirtschaftsinformatik. 46, 331–340 (2004)
16. Fettke, P., Loos, P.: Multiperspective Evaluation of Reference Models – Towards a Framework. In: Jeusfeld, M.A. and Pastor, Ó. (eds.) Conceptual Modeling for Novel Application Domains: ER 2003, Lecture Notes in Computer Science (LNCS). pp. 80–91. Springer, Berlin (2003)
17. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Q. 28, 75–105 (2004)
18. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, xiii–xxiii (2002)
19. Fettke, P.: State-of-the-Art des State-of-the-Art: Eine Untersuchung der Forschungsmethode "Review" innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik. 48, 257–266 (2006)
20. Kitchenham, B., Charters, S.: Guidelines for performing Systematic Literature reviews in Software Engineering. (2007)
21. Becker, J., Delfmann, P., Knackstedt, R.: Adaptive reference modeling: Integrating configurative and generic adaptation techniques for information models. In: Becker, J. and Delfmann, P. (eds.) Reference Modeling Efficient Information Systems Design Through Reuse of Information Models. pp. 27–58. Physica, Heidelberg (2007)
22. Schütte, R., Rotthowe, T.: The Guidelines of Modeling - an approach to enhance the quality in information models Introduction : Theoretical Aspects in Information Modeling. In:

Proceedings of the International Conference on Conceptual Modeling (ER'1998). pp. 240–254. , Singapore (1998)
23. Zwicky, F.: Discovery, invention, research through the morphological approach. MacMillian, New York (1969)
24. Redlein, A.: Fulfilling customer's needs by the use of a variant configurator for dynamic product definition. In: Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'1999). pp. 735–742. , Barcelona (1999)
25. Piller, F., Möslein, K., Stotko, C.: Does mass customization pay? An economic approach to evaluate customer integration. Prod. Plan. Control. 15, 435–444 (2004)
26. Kristianto, Y., Helo, P., Jiao, R.J.: A system level product configurator for engineer-to-order supply chains. Comput. Ind. 72, 82–91 (2015)
27. von Hippel, E., Katz, R.: Shifting Innovation to Users via Toolkits. Manage. Sci. 48, 821–833 (2002)
28. Forza, C., Salvador, F.: Product Information Management for Mass Customization - Connecting Customer, Front-office and Back-office for Fast and Efficient Customization. Palgrave Macmillan, Houndmills (2006)
29. Randall, T., Terwiesch, C., Ulrich, K.T.: Principles for User Design of Customized Products. Calif. Manage. Rev. 47, 68–85 (2005)
30. Füller, K., Böhm, M., Krcmar, H.: Designing for positive user experience in product design: A qualitative analysis of toolkit design elements and their implications on emotional reactions and perceptions. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS'2016). pp. 1810–1819. , Koloa (2016)
31. Roth, M., Nerb, A., Kasperek, D., Lindemann, U.: A Tool to Bridge the Gap from Functional Dependencies to Configuration Rules - Translating Knowledge on Functional Restrictions from Systems Engineers to a Configurator for Sales and Decision Makers. In: Proceedings of the IEEE International Systems Conference (SysCon'2015). pp. 158–163. , Vancouver (2015)
32. Tiihonen, J.: Support for configuration of physical products and services. Aalto University | Doctoral Dissertations, Helsinki (2014)
33. Winter, R.: BEN-Gestaltungsmethodik für Geschäftslösungen. In: Österle, H., Winter, R., and Brenner, W. (eds.) Business Engineering Navigator - Gestaltung und Analyse von Geschäftslösungen "Business-to-IT." pp. 55–84. Springer, Berlin (2011)
34. Yip, A.L.K., Corney, J.R., Jagadeesan, A.P., Qin, Y.: A Product Configurator for Cloud Manufacturing. In: Proceedings of the ASME 2013 International Manufacturing Science and Engineering Conference (MSEC'2013). pp. 1–9. , Madison (2013)
35. Rehse, J.-R., Hake, P., Fettke, P., Loos, P.: Inductive Reference Model Development: Recent Results and Current Challenges. In: INFORMATIK 2016. pp. 739–752. , Klagenfurth (2016)

Terminology for Evolving Design Artifacts

Hannes Schlieter, Jeannette Stark, Martin Burwitz, and Richard Braun

Technische Universität Dresden, Chair of Wirtschaftsinformatik, esp. Systems Development, Dresden, Germany
{hannes.schlieter,jeannette.stark,martin.burwitz,richard.braun}@tu-dresden.de

Abstract. Many design researchers evolve artifacts in succeeding projects. Yet, these researchers lack a terminology to describe how their artifacts evolve. We provide such a terminology by paralleling concepts from evolution with design artifacts using examples from conceptual modeling. We found seven concepts from evolution that we think are useful to describe evolving design artifacts. Evaluating whether these concepts have been addressed, we identified six conceptual modeling design studies, whose authors have addressed some of the concepts with their own words. Using two of these studies, we explain how terminology from evolution can be used to describe evolving design artifacts. We hope that our results are useful to be integrated in design science procedure models to help researchers increasing rigor and relevance of their research, e.g. by allowing to clarify how the artifact at hand has evolved or to describe the evolutionary distance to preceding artifacts.

Keywords: Design artifacts, Evolution, Coevolution, Design Science, Conceptual Modeling.

1 Introduction

Researchers often evolve their modeling artifacts in succeeding design projects. Thereby, artifacts sometimes evolve differently. For example, the Unified Modeling Language (UML) specifications were revised to fix shortcomings and support concept integration (cf. [1]) leading to a successive UML versions. In this type of evolution, successive UML versions evolve within the context of general-purpose modeling and consider e.g. changed user needs. An example for another type of evolution comprises parts of the UML being transferred to the field of systems engineering, leading to SysML [2]. In this case evolution happens within a specific domain so that design researchers need to address the specific requirements of the new domain. Yet another type of evolution may be distinguished for versions of UML (e.g. UML 2.0 and 2.5) and corresponding revisions of standards-compliant UML modeling tools (e.g. [3]) that were evolved based on revisions of the UML. In this type of evolution, evolving the modeling tool depends on how the corresponding modeling language evolves and is thus, dependent on the evolution of another artifact.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Prior design research has discussed evolution in the context of mutability [4–6]. Mutability is a core component of design research and researchers are asked to address mutability when publishing their results [7]. For mutability, two paradigmatic perspectives are distinguished: in-design and in-use [4, 5]. In-design mutability refers to the degree of artifact change encompassed by the design theory [7] and allows for designing mutable artifacts that are adaptable to varying organizational contexts [4] and situations [5]. In-use mutability refers to artifacts that evolve over time with users adapting the artifact to new situations, e.g. through configuration [5] or tailoring [8]. Gregor and Jones state that mutability goes even further by "allowing for a certain amount of adaptation or evolution", [7], p. 326. However, how artifacts evolve and adapt is so far not specified in in-use and in-design mutability, so that design researchers lack a specification to describe artifact evolution and hence, cannot address mutability of their design artifacts for this particular aspect.

Specifying how artifacts evolve may help design researchers to increase rigor and relevance for their design projects. Rigor in design science depends on how researchers ground their project with existing artifacts [9]. Being able to specify how artifacts evolve instead of just claiming that it evolves may help identifying important requirements for the evolution at hand. Relevance in design science depends on how researchers address opportunities and problems with their artifacts in the application domain [9, 10]. Specifying how their artifact evolves from preceding artifacts may help to address relevance more precisely. For example, differentiating between different types of evolution allows for discussing the evolutionary distance more precisely e.g. by explaining what features have been added or removed, whether new domain has been entered or an artifact has now coevolved with another artifact. Beyond helping researchers to enhance rigor and relevance, specifying how artifacts evolve contribute to discuss evolutionary lineages on different levels such as single artifacts or artifacts of a domain such as conceptual modeling design artifacts. Investigating evolutionary lineages for single artifacts can also help to discuss and compare maturity of different artifacts and to reveal future potential for evolution. Investigating evolutionary lineages of different domains contribute also to the discussion of how mature domains, such as conceptual modeling, are in contrast to other IS domains. Furthermore, discussing evolution in the context of mutability may help to further advance mutability research.

In this paper we aim to investigate concepts to specify evolving design artifacts. Although we believe that specifying design artifacts is important for all IS domains that develop design artifacts we focus on conceptual modeling design research in this paper. Conceptual modeling is a major topic of the Wirtschaftsinformatik domain, which is primarily design-oriented [11]. Furthermore, this domain was chosen as the authors are familiar with it so that concepts can be transferred to this domain more easily and better examples can be given than for other domains. Finally, to allow for comparing evolutionary lineages between domains, other domains will follow in future research. This paper represents an interim struggle in theorizing [12] that needs to be built on to provide for more rigor and relevance in design science projects, to develop and compare evolutionary lineages and to enhance mutability research.

Hence, these research goals still need to be addressed in succeeding research. Yet, this paper also contributes a specific outcome and concepts discussed in this paper can already be used as a terminology to describe evolution more specific than it was possible before so that evolution as one aspect of mutability can already be specified more precisely as demanded by [7]. For investigating concepts to describe evolution and deriving the terminology of evolving design artifacts, we use argumentative reasoning between evolution of design artifacts and evolution in biology. For a first proof of relevance, we employ a content analysis according to Krippendorff [13]. Using this method, we investigate whether concepts identified to describe evolution have already been applied in recent conceptual modeling design science research.

2 Theoretical Background

2.1 Mechanisms that Drive Evolution

Evolution is described as a change in the heritable characteristics of populations over successive generations [14] and is pushed forward through diverse mechanisms, including mutation, migration, natural selection and gene drift [15]. **Mutations** alter the genome of organisms and result from random errors during DNA replication [16]. For example, a mutation can cause a change in a beetle's genotype from genes for grey to white coloration (Fig. 1a). Those beetles that gain genes for white coloration can have offspring with a gene for white coloration, too. Accordingly, genes for white coloration can be more frequent in future populations. Mutations may result in a gain of a new feature or a loss of an ancestral feature. For example, the bacterium Sphingobium has evolved a metabolic pathway as a new feature to degrade toxins [17] whereas birds have lost their wing function when there was no need to fly. In analogy to evolution in biology, we refer to mutation of design artifacts as a gain of a new function or withdrawal of an ancestral function. Yet, there are also differences within these two contexts. While in biology mutations occur randomly and might have a positive, neutral or negative impact on survival and reproduction capabilities [16], mutations in design science can typically be planned and can hence, rather be classified as non-random. As design researchers plan mutations also the impact is more likely to be positive than in biology. Nonetheless, using the analogy in conceptual modeling design science may help to specify a gain or a loss of an ancestral feature. Yet, the implication of impacting randomly versus non-randomly needs further investigation for design artifacts and should therefore not be made. As an example, for conceptual modeling design artifacts, we refer to the Business Process Modeling and Notation (BPMN) that has experienced major and minor revisions that have continuously produced a variety of BPMN features. For example, the major revision that lead to version 2.0, not least introduced the choreography diagram as a mean to model interactions of multiple communication partners. An additional feature of this mutation comprises the extension mechanism that enables modelers to systematically extend the BPMN core by custom elements.

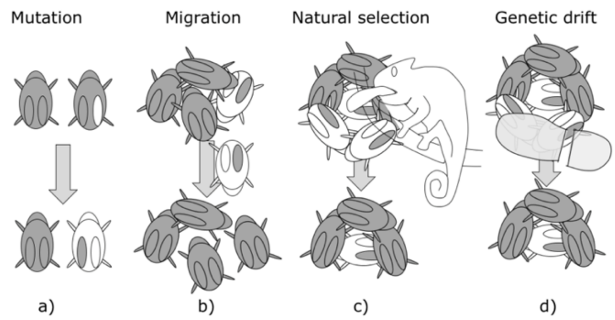


Figure 1. Evolutionary mechanisms: mutation and migration to induce variation as well as natural selection (non-random) and genetic drift (random) for transferring variation

While mutations lead to genetic variation within a population, **migration** allows transferring this variation to further populations of a specie (Fig. 1b). Imagine, beetles with genes for white coloration migrate into another population and thus, allow for variety in a second population. Please note that a specie usually includes several populations. A population is defined as a group of organisms, in which reproduction takes place [18]. In contrast, organisms of a specie only potentially interbreed as they live in distinct areas and thus, may form separate populations whose organisms do usually not interbreed. Migration allows referring to transferring design artifacts to another context and thus, give way for an own evolutionary lineage. For example, BPMN has initially addressed the management of business processes, but has already been used in a variety of other domains, that even sometimes lead to new BPMN species by using the extension mechanism (e.g. [19–22]).

Mutation and migration allow for variation within populations and create the basis for two further mechanisms: natural selection and gene drift. **Natural selection** acts on variation by non-random sorting [23]. Imagine, white beetles live in a habitat with green plants so that predators can easily spot them. In this case, they are not likely to survive and reproduce, which lead genes for white coloration being reduced non-randomly for future populations (Fig. 1c). Also, **gene drift** acts on variation. Yet, while natural selection acts non-randomly, gene drift acts randomly [24]. For example, a beetle population with white and grey coloration is threatened by a storm, in which most of the white beetle die resulting in a random drift towards grey beetles (Fig. 1d). Genetic drift may cause two populations starting with the same genotype drifting apart into divergent populations or even species [25].

2.2 Outcome of Evolution

Evolutionary mechanisms lead to diverse evolutionary outcomes such as speciation [26], adaptation [27, 28], coevolution [29], and extinction [30]. **Speciation** refers to populations of one specie diverging into two or more species and is made responsible for diversity of life [31]. After having diverged, organisms of the two species might still produce common offspring that are in most cases infertile such as mules resulting from horses reproducing with donkeys [32]. A specie is used as a concept to differentiate organisms that cannot reproduce even if manipulated (e.g. moved to another area) while population is used to differentiate organisms of a specie that actually do not mate, e.g. as they live in different areas or times. In design science, **species** may be used in analogy to **artifact types** as this concept helps indicating that design artifacts can potentially be used to produce offspring that address similar requirements or allow to solve similar tasks. In this sense, succeeding artifacts do not represent distinct artifact types but rather succeeding artifact instances of the same type. Furthermore, **populations** may be used in analogy to **artifact variants** to indicate that instances of artifact types can be migrated into another context. For example, the core BPMN 2.0 can be considered a BPMN artifact type, which meets requirements for modeling business processes. Using the core BPMN 2.0 for software process modeling [20] led to a new artifact variant that was able to meet the same requirements in a new context. Based on this artifact variant, the revision and extension of the core BPMN 2.0 regarding new context-related requirements evolved a new BPMNt artifact type. While **speciation** leads to a new specie, **extinction** means that an entire specie disappears. In biology, speciation and distinction happen regularly [30]. When using speciation and extinction to describe design artifacts, we might rather refer to **entry** and **exit** in analogy to the product life cycle [33].

Adaptation refers to organisms evolving adaptive traits by mutation and natural selection and characterizes a population's movement towards a fit to the environment [34]. Adaptations mostly occur through mutations, which impact positively, negatively or neutrally on the organism's survival and reproduction capabilities. Based on how mutations impact, they may prevail or be sorted out. Adapting artifact instances by introducing a new function usually occurs in succeeding **artifact versions**. For example, the introduction of the choreography diagram as a new feature in version BPMN 2.0 can be considered an adaptation to the rising demand for modeling and managing communication between business partners (cf. [35]).

Coevolution occurs across pairs of species that interact with each other [36]. These interactions can lead to adaptations of one specie to cause adaptations in another specie and vice versa. Accordingly, a cycle of adaptations emerges that is referred to as coevolution [29]. Based on how species interact, coevolution is distinguished into conflict and cooperation. Conflict usually emerges between predators and prey. Predators can coevolve to catch their prey more effectively. As a result, selective pressure increases for their prey and leads to coevolving the prey, e.g. to escape more effectively. Adaptive traits developed in conflict comprise speed of movement, ability to detect and habitat choice [37]. Cooperation e.g. occurs between insects and flowers that reward insects with nectar but rely on them for reproduction. Insects and flowers have evolved interactions such as communication via scent or attraction via color [38]. Coevolution has already been applied to organizational adaptation [39], strategic management [40], and strategic alliances [41]. It has also been used for IS by discussing mutual adaptations of business and IS strategies and for coevolutionary choices of platform-based ecosystem owners [42] and the ecosystem's environment [43]. Using coevolution as analogy for evolving design artifacts, we consider the BPMN an artifact type in coevolution with process execution languages (e.g. XPDL or BPEL) that represent business processes in context of machine-based workflow execution. These artifact types and the BPMN artifact type complement one another in terms of language concepts, semantics and interchange formats [63].

Coevolution in biology may lead to **endosymbiosis** that references to cases where one partner is integrated into the other partner's body [44]. This concept has been shaped by Mereschkowsky [45] and has been used as example that evolution may lead to more complexity. Endosymbiosis can be used as analogy to describe evolving design artifacts that incorporate versions of other artifact types. Referring the example of BPMN, the *Lane* concept may be classified as endosymbiotic integration of swim lanes from other modeling approaches (cf. [46, 47]). Please note a distinction between endosymbiosis and coevolution: In coevolution an artifact type adapts with another artifact type. In case of endosymbiosis only certain versions of an artifact type may be incorporated while other versions of that artifact type may remain independent.

3 Concepts and Relationships for Evolving Design Artifacts

We summarize concepts and relationships that connect these concepts in Table 1. Furthermore, we show how these concepts and relationships interrelate in Fig. 2. Examples are taken from conceptual modeling languages. This provides a larger scope of consideration in comparison to the specific BPMN examples in Section 2. As depicted in Fig. 2, we parallel species with artifact types, in analogy to an instance-type relationship, to reflect the idea behind its instances. The artifact instances, which in biology reflect organisms, can further be paralleled with artifact versions indicating that versions evolve, e.g. during iterations [48], [9]. Furthermore, populations are paralleled with artifact variants as this concept allows for summarizing versions that are designed for different application domains and that can develop their own evolutionary lineage. Among relationships, adaptation refers to gaining new or loosing ancestral functions that lead to a better fit to the application domain. Accordingly, adaptations reflect iteration cycles in design science.

Migration relates artifact versions to variants indicating that a particular version is used for another application domain. A further relationship, that we found useful to describe evolving design artifacts is coevolution. Coevolution can happen between versions of one artifact type and those versions of another artifact type and is therefore used as relationship between artifact types. In contrast, endosymbiosis may be used with the term incorporation for design artifacts, meaning that artifact versions of one artifact type may be incorporated for artifact versions of another artifact type.

Table 1. Notions to describe evolving design artifacts and examples from conceptual modeling

Concepts from biology	***Use for IS design artifacts***	***Example***
Specie: Potentially interbreeding organisms	Artifact type: Summarizing artifacts with similar goals and user requirements	Enterprise Modeling Languages
Population: Organisms of a specie that actually mate, e.g. as they live in the same habitat	Artifact variant: Summarizing artifact versions that are used for the same application domains	Enterprise Architecture Languages, Business Process Modeling Languages, Software Modeling Languages
Organism: Instance of a specie	Artifact version: Instance of an artifact type	BPMN 1.0/1.1/1.2/2.0 [49], UML 1.0-1.5/2.0/2.5 [1]
Relationship from biology		
Adaptation: Developing or losing traits as well as non-random sorting towards a fit to the habitat	Adaptation: developing or losing functions and sorting these functions, e.g. towards user acceptance	Revisions of UML specifications in order to fix shortcomings and support concept integration, for instance (cf. [1])
Migration: Transferring variation (e.g. mutations) in between populations	Migration: e.g. using an artifact version within another context	Parts of UML are transferred to the field of systems engineering, leading to SysML [2]. Some parts are completely reused (e.g., Use Case Diagrams), while other parts are adapted domain-specifically (e.g. Block Definition Diagrams [50])
Coevolution: Evolution of a species causes adaptations in a second specie and vice versa. Is distinguished in conflict and cooperation.	Coevolving in cooperation: Artifact types that evolve together (modeling grammar and its tool)	Versions of the UML (e.g. UML 2.0 and 2.5) and corresponding revisions of standards-compliant UML modeling tools (e.g. [3])
Endosymbiosis: Incorporation of an organism within another species' organisms	Incorporation: Artifact versions of an artifact type being incorporated in versions of another artifact type	Use Case Diagrams were integrated in UML (cf. [51]). Parts of the UML were refined in SysML (e.g. Ports [2])

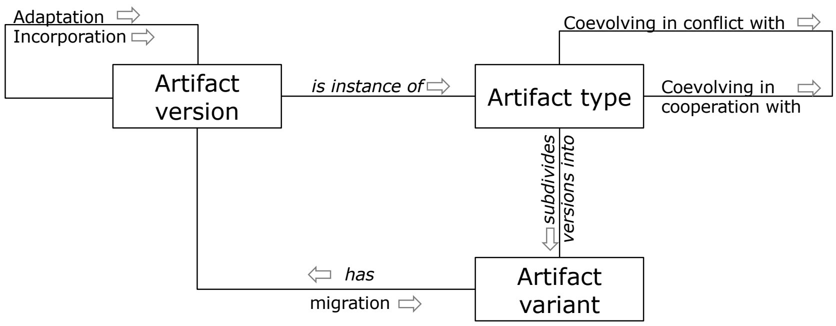


Figure 2. Relating concepts and relationships derived from evolutionary biology (cursive relationships are only integrated for a better readability)

4 Evaluation

Although this works focusses on discussing concepts and relationships that might be used for further theorizing in future work, we include a pilot content analysis following Krippendorff [13] to investigate whether these concepts and relationships are relevant to describe evolving design artifacts. We believe that “when theories are particularly interesting or important, there should be greater leeway in terms of empirical support” [52], p. 383, allowing for just a small demonstration to show that concepts and relationships of this study may be of relevance. Of course, subsequent research is required to assess whether concepts and relationship will hold up [52]. In this pilot study we focus on recent conceptual modeling design science research and limit the analysis to articles published in the Business & Information Systems Engineering (BISE)-journal during the last five years. This journal represents the flagship of the Wirtschaftsinformatik domain, which is primarily design-oriented [11].

We found six conceptual modeling design studies that we have summarized in Table 2. In these studies, we could detect all relationships identified in Section 2 and 3, including adaptation, migration, incorporation and coevolution. Hence, we could provide a first proof that identified concepts and relationships are relevant to describe evolving conceptual modeling design artifacts. Furthermore, we identified the words the authors used to refer to these concepts. We found that adaptation is primarily referred to with “extend” while migration is rather referred to “extent” plus a specification such as “extends by combining” [53], p. 65 or “extends with the required aspects” [54], p. 318. In contrast, incorporation was referred to in different ways using words such as “building on” [55], p. 252, and “the techniques were sequentially arranged to form” [55], p. 258. Likewise, coevolution was referred to in different ways such as with “corresponding software prototype” [55], p. 251. Although we found words how authors describe the different types of evolution that we have presented in Table 1, we refrain from suggesting words to address these types of evolution. For discussing how authors may refer to these concepts we expect that more empirical research is required.

Providing an example how to use the terminology, we use the study from Johannsen and Fill [55] in Fig. 3a and Gailly et al. [59] in Fig. 3b. Johannsen and Fill have developed an artifact type that represents a roadmap for Business Process Improvement (BPI) with conceptual modeling support. Whereas in their first version they describe the method [61], they add conceptual modeling support in succeeding studies [55], [62]. The authors describe these artifact versions as first contribution (BPI-roadmap) and second contribution (modeling support) [55], p.252 and develop these contributions in separate papers. Johannsen and Fill further incorporate versions of artifact types such as SIPOC and Ishikawa, which they further describe as foundations of their artifact type [55], p.253f. Furthermore, we found their artifact type in coevolution with the ADOxx platform [63]. This platform has been developed by members of the current research group so that implications of the evolution of one artifact may be implemented for the other artifact.

Table 2. Studies that evolve conceptual modeling design artifacts

Ref.	*Artifact type*	*Types of evolution*	*Coding examples*
[56]	Business Activities for modelling Processes and process-rel. RBAC models	**Adaptation** of Business Activities UML 2.0 extension [57] for process-related delegation models as a new artifact version	"[…] implemented our approach as a delegation extension […]", p. 235
[53]	Business model repr. lang. for business model analysis	**Migration** of VDML [58] by adaptation for specialized VDML business model constructs and views	"This paper extends the previous work by combining […]", p. 65
[59]	Recommendation-based conceptual modeling and ontology evolution framework (CMOE+)	**Incorporation:** The variant incorporates e.g. versions of the of the Unified Foundational Ontology [60] **Migration**: Conceptualization of a generic framework (CMOE+) and extending this framework for a BPMN instantiation as a new variant **Coevolution**: The variant exists in coevolution with the eclipse plugin extension for BPMN2 Modeler	"[UFO] was selected as a core ontology", p. 242 "CMOE+ […] may be instantiated and further specialized to support different concrete modeling languages.", p. 236
[55]	Business process improvement (BPI) roadmap with conceptual modeling support	**Incorporation**: the BPI roadmap (first version) incorporates versions of artifact types such as SIPOC and KPI **Adaptation**: Conceptualizing a BPI roadmap [61] and adapting it for further conceptual modeling support [62], [55] as a new artifact version. **Coevolution** with ADOxx [63]	"The roadmap […] is […] building on BPI techniques", p. 252; "the techniques were sequentially arranged to form a roadmap", p. 258 "Based on this BPI roadmap, a domain-specific conceptual modeling method (DSMM) has been developed.", p. 251 "a corresponding software prototype has been implemented using a meta modeling platform ", p. 251
[54]	Extended ITML for IT platforms and multilevel model of IT platforms	**Migration** of Memo ITML [64] for IT platform constructs as a new variant of MML [65] **Coevolution**: Memo ITML for IT platforms exists in coevolution with the Flexible Meta modeling language implemented in Xmodeler [66]	"we undertake an attempt to extend ITML […] with the required aspects", p. 318 "we apply FFML to model different aspects of IT platforms.", p. 234
[67]	Business-oriented Service Description Language	**Migration**: Developing a UML-profile for the graphical Notation of BoSDL as a variant of UML	"[…] develop such a format based on the UML, which we adapted by creating a profile", p. 12

Gailly et al. [59] developed a recommendation-based conceptual modeling and ontology evolution framework (CMOE+) that they present as a generic framework in its first version. In its second version a specific instantiation for CMOE+BPMN was developed. Because of the different versions being either generic (CMOE+) or specific (CMOE+BPMN) we present these versions as variants. The authors describe that this artifact type "might further be instantiated and further specialized to support different concrete modeling languages" [59], p.236 indicating that there may be further variants. Gailly et al. incorporate versions of artifact types such as the Unified Foundational Ontology (UFO) [60] into their artifact. Furthermore, they describe that they implemented an eclipse plugin in their BPMN2 modeling tool, which may indicate that CMOE+BPMN is in coevolution with BPMN2 modeler. Please note that these two examples represent our interpretations of the authors' publications. Although two authors of this paper coded independently and discussed and merged their results, our view may include misinterpretations and is limited to aspects we found relevant. We suggest the terminology for evolving design artifacts to be more powerful in case the authors use it themselves to describe their artifact evolution rather than other researchers trying to interpret their results.

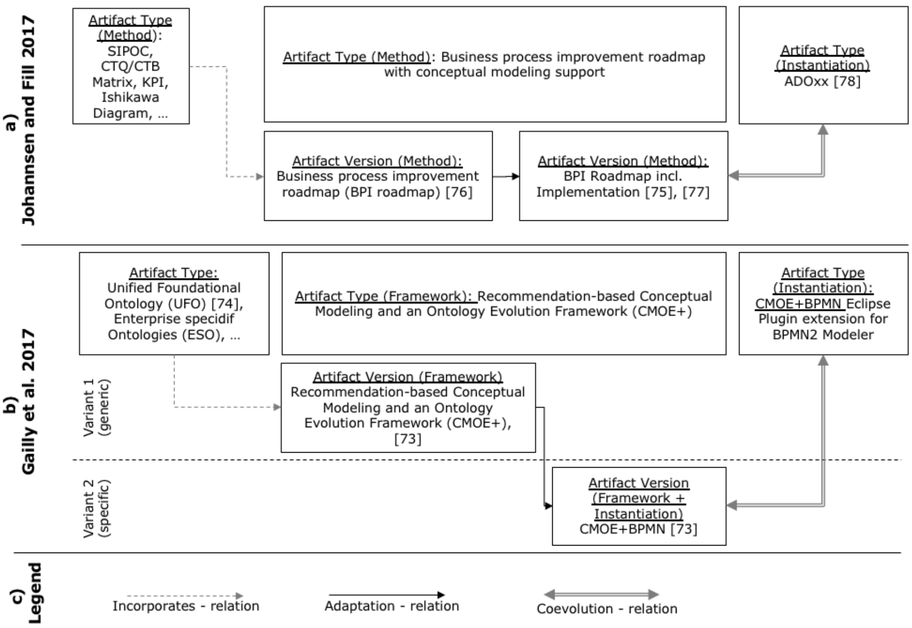


Figure 3. Describing an artifact evolution using a) the example from [55], b) the example from [59] and c) providing a legend for the different relations that occur in evolving design artifacts

5 Discussion

The terminology to describe evolving design artifacts, summarized in Table 1 and Fig. 2, can be used to help researchers to increase rigor by grounding their projects with existing artifacts. Accordingly, design researchers may clarify how present and prior artifacts relate to one another or can describe artifacts that they coevolve with. Thereby, design researchers address mutability of their design artifacts such as Gregor and Jones asked for [7]. Furthermore, we suggest that the terminology for evolving design artifacts helps to decide where to focus on when evaluating design artifacts as suggested by Gregor and Hevner [68]. For example, when adapting an artifact version that has been evaluated profoundly and that has positive impact, evaluations of the new version may focus on those features that have been added or abandoned instead of again evaluating the whole version. In contrast, when evaluating an artifact variant, evaluation can focus on distinct characteristics of the application domain the variant is used in.

To improve relevance, design researchers can use the terminology for evolving design artifacts to clarify their contribution. In analogy to the design science contribution types presented in [68], design researchers can distinguish different contribution types for their artifacts. As one contribution type, their artifact instance may base on well-developed and evaluated artifact instances and thus, include mature knowledge. In this case, design researchers could discuss the distance between their current and preceding artifact instances, e.g. by devising the number of new functions or abandoned functions, as well as the effort required to implement or abandon these functions. As another contribution type, researchers might describe their design artifacts to only have a limited evolutionary lineage. In this case, design researchers justify their contribution by discussing the novelty and originality of their ideas as well as the problems their artifacts help to solve. Furthermore, as mutations in design projects are subject to human activity, relevance for future design projects might be increased by facilitating mutations. For example, allowing for more creativity in design projects or for appropriate time slots for evolving the design artifact might impact on the number and quality of mutations. For instance, the act of creativity has been discussed as bisociation, that can be achieved by relating issues that at first seem unrelated to produce new insights [69]. Bisociation has been discussed in psychology but to our knowledge not been reflected for design science.

So far, the terminology for evolving design artifacts does not fully relate to recent mutability approaches. In-design mutability refers to the degree of change encompassed within one artifact version [7] and does not reflect relationships between artifacts. In-use mutability describes artifacts to evolve over time by allowing users to adapt the artifact. However, also designers evolve their artifacts (e.g. [53–55, 59]), which is not reflected in in-use mutability. We suggest that the terminology for evolving design artifacts may enhance in-use or may be treated as a unique aspect.

6 Limitations

As this work represents an interim struggle in theorizing we include a discussion of potential boundaries that need further investigation. We have limited our analysis to conceptual modeling design research but presume that these concepts can be applied to other IS domains that deal with design science. We suspect that coevolution is possibly not a relevant type to describe artifact evolution in any IS domain. In conceptual modeling, coevolution was reconstructed for artifact types and their corresponding modeling tool. Accordingly, for this domain we found coevolution to be tool-dependent and might hence not apply to any domain. Yet, more research is needed to specify whether coevolution is indeed dependent on tools.

Furthermore, one might question whether concepts of the terminology for evolving design artifacts may apply to any conceptual modeling project. In the evaluation we have only focused on types of artifact evolutions that were published in the BISE-journal from 2014-2018. We found that artifact evolutions had a low to mid-range complexity including one to three different evolutionary relationships. For example, Fill and Johannsen described an artifact evolution that includes an incorporation, adaptation as well as coevolution [55]. In contrast, other artifact evolutions only included one type of artifact evolution and might in this sense be described as less complex in terms of evolutionary steps. Yet, we suspect that concepts of the terminology for evolving design artifacts may be also used for more complex evolutions such as the evolution of the UML or BPMN. We have already pinpointed to how these different types of evolution may correspond to the development of UML in the introduction but agree that more research is needed to proof whether concepts can also be applied to large artifact design projects.

7 Conclusion and Future Research

As design artifacts evolve over time, design researchers need a terminology to describe this evolution. We have developed such a terminology based on evolution in biology and evaluated whether its concepts and relationships are already used, although in other words. We found that authors describe evolution in their own words but lack a consistent terminology. For further evaluation we suggest to use a larger sample size and to check the coders' interpretations against the designers' experience by conducting interviews. Thereby, usefulness of the terminology and the need for further notions may be assessed. So far, we only provide examples of evolving artifacts [55, 59]. A systematic approach to describe evolving design artifacts might help to compare evolutions of different design artifacts and thus, to render evolution of design artifacts more transparent. The terminology derived in this paper bridges the gap between the need to address mutability [7] and the guidance required to do so [70]. So far, this guidance is available as a terminology and not as a mid-range theory [71] between evolution in biology and evolution in design artifacts. To further approach theory, e.g. boundaries need to be investigated, concepts need to be refined for other domains, propositions about e.g. typical sequences in which types of evolutions occur should be developed and translated into hypotheses.

References

1. Platt, R., Thompson, N.: The Evolution of UML. In: Encyclopedia of Information Science and Technology, Third Edition. pp. 1931–1936. IGI Global (2015).
2. Object Management Group: Systems Modeling Language (SysML) - Verion 1.5, (2017).
3. Modelio, https://www.modelio.org/. Download: 2018-08-01
4. Pöppelbuss, J., Goeken, M.: Understanding the Elusive Black Box of Artifact Mutability. In: Wirtschaftsinformatik. pp. 1557–1571 (2015).
5. Sjöström, J., Agerfalk, P.J., Lochan, A.R.: Mutability Matters: Baselining the Consequences of Design. Procedings MCIS. (2011).
6. Gregor, S., Iivari, J.: Designing for mutability in information systems artifacts. Inf. Syst. Found. Theory Represent. Real. Eds Hart Gregor Aust. Natl. Univ. Press Canberra (2007).
7. Gregor, S., Jones, D.: The anatomy of a design theory. J. Assoc. Inf. Syst. 8, 312–335 (2007).
8. Trigg, R.H., Bødker, S.: From implementation to design: tailoring and the emergence of systematization in CSCW. In: Conf. on Comp. supp. coop. work. pp. 45–54. ACM (1994).
9. Hevner, A.R.: A three cycle view of design science research. Scand. J. Inf. Syst. 19, 4 (2007).
10. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 24, 45–77 (2007).
11. Winter, R.: Design science research in Europe. Eur. J. Inf. Syst. 17, 470–475 (2008).
12. Weick, K.E.: What theory is not, theorizing is. Adm. Sci. Q. 40, 385–390 (1995).
13. Krippendorff, K.: Content analysis: An introduction to its methodology. Sage (2012).
14. Hall, B., Strickberger, M.W.: Strickberger's evolution. Jones & Bartlett Learning (2008).
15. Hartl, D.L., Clark, A.G., Clark, A.G.: Principles of population genetics. Sinauer associates Sunderland (1997).
16. Mayr, E.: What evolution is. Science Masters Series (2001).
17. Copley, S.D.: Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends Biochem. Sci. 25, 261–265 (2000).
18. Dobzhansky, T., Dobzhansky, T.G.: Genetics of the evolutionary process. Columbia University Press (1970).
19. Kopp, O., Binz, T., Breitenbücher, U., Leymann, F.: BPMN4TOSCA: A Domain-Specific Language to Model Management Plans for Composite Applications. In: Mendling, J. and Weidlich, M. (eds.) Business Process Model and Notation. pp. 38–52. Springer (2012).
20. Pillat, R.M., Oliveira, T.C., Alencar, P.S.C., Cowan, D.D.: BPMNt: A BPMN extension for specifying software process tailoring. Inf. Softw. Technol. 57, 95–115 (2015).
21. Graja, I., Kallel, S., Guermouche, N., Kacem, A.H.: BPMN4CPS: A BPMN Extension for Modeling Cyber-Physical Systems. In: 2016 IEEE 25th Int. Conf. on Enabling Techn.: Infrastr. for Coll. Ent. pp. 152–157 (2016).
22. Yousfi, A., Bauer, C., Saidi, R., Dey, A.K.: uBPMN: A BPMN extension for modeling ubiquitous business processes. Inf. Softw. Technol. 74, 55–68 (2016).
23. Gregory, T.R.: Understanding natural selection: essential concepts and common misconceptions. Evol. Educ. Outreach. 2, 156 (2009).
24. Masel, J.: Genetic drift. Curr. Biol. 21, R837–R838 (2011).
25. Lande, R.: Fisherian and Wrightian theories of speciation. Genome. 31, 221–227 (1989).

26. Gavrilets, S.: Perspective: models of speciation: what have we learned in 40 years? Evolution. 57, 2197–2215 (2003).
27. Dobzhansky, T.: On some fundamental concepts of Darwinian biology. In: Evolutionary biology. pp. 1–34. Springer (1968).
28. Nakajima, A., Sugimoto, Y., Yoneyama, H., Nakae, T.: High-Level Fluoroquinolone Resistance in Pseudomonas aeruginosa Due to Interplay of the MexAB-OprM Efflux Pump and the DNA Gyrase Mutation. Microbiol. Immunol. 46, 391–395 (2002).
29. Wade, M.J.: The co-evolutionary genetics of ecological communities. Nat. Rev. Genet. 8, 185 (2007).
30. Benton, M.J.: Diversification and extinction in the history of life. Science. 268, 52–58 (1995).
31. Wilson, R.A.: Species: New interdisciplinary essays. (1999).
32. Short, R.V.: The contribution of the mule to scientific thought. J. Reprod. Fertil. Suppl. 359–364 (1975).
33. Klepper, S.: Entry, exit, growth, and innovation over the product life cycle. Am. Econ. Rev. 562–583 (1996).
34. Orr, H.A.: The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6, 119 (2005).
35. Decker, G., Kopp, O., Leymann, F., Pfitzner, K., Weske, M.: Modeling service choreographies using BPMN and BPEL4Chor. In: Int. conf. on Adv. Inf. Syst. Eng. pp. 79–93. (2008).
36. Futuyma, D.J., Slatkin, M.: Coevolution. Sinauer Associates Incorporated (1983).
37. Sih, A.: Predator and prey lifestyles: an evolutionary and ecological overview. Predation Direct Indirect Impacts Aquat. Communities. 203–224 (1987).
38. Lunau, K.: Adaptive radiation and coevolution—pollination biology case studies. Org. Divers. Evol. 4, 207–224 (2004).
39. McKelvey, B.: Perspective-quasi-natural organization science. Organ. Sci. 8, 351–380 (1997).
40. Barnett, W.P., Hansen, M.T.: The red queen in organizational evolution. Strateg. Manag. J. 17, 139–157 (1996).
41. Koza, M.P., Lewin, A.Y.: The co-evolution of strategic alliances. Organ. Sci. 9, 255–264 (1998).
42. Peppard, J., Breu, K.: Beyond alignment: a coevolutionary view of the information systems strategy process. ICIS 2003 Proc. 61 (2003).
43. Tiwana, A., Konsynski, B., Bush, A.A.: Research commentary-Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. Inf. Syst. Res. 21, 675–687 (2010).
44. Kutschera, U., Niklas, K.J.: Endosymbiosis, cell evolution, and speciation. Theory Biosci. 124, 1–24 (2005).
45. Mereschkowsky, C.: Über natur und ursprung der chromatophoren im pflanzenreiche. Biol. Cent. 25, 293–604 (1905).
46. Rummler, G.A., Brache, A.P.: Improving performance: How to manage the white space on the organization chart. John Wiley & Sons (2012).
47. Binner, H.F.: Anforderungsgerechte Datenermittlung für Fertigungssteuerungssysteme. Beuth Verlag, Berlin (1987).
48. Verschuren, P., Hartog, R.: Evaluation in design-oriented research. Qual. Quant. 39, 733–762 (2005).
49. Chinosi, M., Trombetta, A.: BPMN: An introduction to the standard. Comput. Stand. Interfaces. 34, 124–134 (2012).

50. Weilkiens, T.: Systems engineering with SysML/UML: modeling, analysis, design. Elsevier (2011).
51. Jacobson, I.: Object-oriented software engineering: a use case driven approach. Pearson Education India (1993).
52. Sutton, R.I., Staw, B.M.: What theory is not. Adm. Sci. Q. 371–384 (1995).
53. Roelens, B., Poels, G.: The development and experimental evaluation of a focused business model representation. Bus. Inf. Syst. Eng. 57, 61–71 (2015).
54. Kaczmarek-Hess, M., de Kinderen, S.: A multilevel model of IT platforms for the needs of enterprise IT landscape analyses. Bus. Inf. Syst. Eng. 59, 315–329 (2017).
55. Johannsen, F., Fill, H.-G.: Meta Modeling for Business Process Improvement. Bus. Inf. Syst. Eng. 59, 251–275 (2017).
56. Schefer-Wenzl, S., Strembeck, M.: Modeling Support for Role-Based Delegation in Process-Aware Information Systems. Bus. Inf. Syst. Eng. 6, 215–237 (2014).
57. Strembeck, M., Mendling, J.: Modeling process-related RBAC models with extended UML activity models. Inf. Softw. Technol. 53, 456–483 (2011).
58. OMG: Value Delivery Modeling Language (VDML), (2014).
59. Gailly, F., Alkhaldi, N., Casteleyn, S., Verbeke, W.: Recommendation-Based Conceptual Modeling and Ontology Evolution Framework (CMOE+). Bus. Inf. Syst. Eng. 59, 235–250 (2017).
60. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.: Towards ontological foundations for conceptual modeling: the unified foundational ontology (UFO) story. Appl. Ontol. 10, 259–271 (2015).
61. Johannsen, F., Fill, H.-G.: Codification of knowledge in business process improvement projects. (2014).
62. Johannsen, F., Fill, H.-G.: Supporting business process improvement through a modeling tool. In: Domain-Specific Conceptual Modeling. pp. 217–237. Springer (2016).
63. Fill, H.-G., Karagiannis, D.: On the conceptualisation of modelling methods using the ADOxx meta modelling platform. Enterp. Model. Inf. Syst. Archit. 8, 4–25 (2013).
64. Heise, D.: Unternehmensmodell-basiertes IT-Kostenmanagement als Bestandteil eines integrativen IT-Controllings. Logos Verlag Berlin GmbH (2013).
65. Frank, U.: The MEMO meta modelling language (MML) and language architecture. ICB-research report (2011).
66. Frank, U.: Multilevel modeling. Bus. Inf. Syst. Eng. 6, 319–337 (2014).
67. Schlauderer, S., Overhage, S.: BoSDL: An Approach to Describe the Business Logic of Software Services in Domain-Specific Terms. Bus. Inf. Syst. Eng. 1–21.
68. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Q. 37, (2013).
69. Koestler, A.: The act of creation. (1964).
70. Gehlert, A., Schermann, M., Pohl, K., Krcmar, H.: Towards a Research Method for Theory-driven Design Research. In: Wirtschaftsinformatik (1). pp. 441–450 (2009).
71. Kuechler, W., Vaishnavi, V.: A framework for theory development in design science research: multiple perspectives. J. Assoc. Inf. Syst. 13, 395 (2012).

Business Role-Object Specification: A Language for Behavior-aware Structural Modeling of Business Objects

Hendrik Schön[1], Susanne Strahringer[1], Frank J. Furrer[2], and Thomas Kühn[2]

[1] TU Dresden, Business Informatics, esp. IS in Trade and Industry, Dresden, Germany
{hendrik.schoen,susanne.strahringer}@tu-dresden.de
[2] TU Dresden, Software Technology Group, Dresden, Germany
frank.j.furrer@bluewin.ch, thomas.kuehn3@tu-dresden.de

Abstract. Representing and reusing the business objects of a domain model for various use cases can be difficult. Especially, if the domain model is acting as a template or a guideline, it is necessary to map the enterprise's individual structure and processes on the shared domain model. Structural modeling languages often do not meet this requirement of reusing structures and complying to established processes. We propose a modeling language called BROS (Business Role-Object Specification) for describing the business objects' structure and behavior for structural models, based on a given domain model and process models. It utilizes roles for a use case related specification of business objects as well as events as interfaces for the business processes affecting these roles. Thus, we are able to represent and adapt the business object in different contexts with individual requirements, without changing the underlying domain model. We demonstrate our approach by modeling a simple case.

Keywords: Structural Model, Role, Behavior, Business Logic

1 Introduction

Software modeling enables the design, construction, configuration, and checking of static and dynamic parts of a software system, that is the structure and behavior, before implementation. Especially for enterprise information systems, we are in need of modeling new or changed model parts, as enterprises frequently use individual, enterprise-specific business processes. Therefore, during modeling of information systems, one often encounters conflicting goals and has to choose between either following a unified and possibly standardized domain model or adhering to the individual and unique business logic of the targeted system context. On the one hand, software engineers want to stick to the principle of standardization and rely on the unified domain model. On the other hand, stakeholders want their individual requirements to be implemented.

When modeling a software system, the model should have four characteristics: clarity, commitment, communication, and control [1]. Especially the last two aspects are important: "The model truly and sufficiently represents the key properties of the real world to be mapped into IT solutions", and "The model is used for the assessment

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

of specifications, design, implementation, reviews and evolution." [1] Both aspects need to be fulfilled if the model is to be effectively used in software construction. However, current models for such software construction are strong in modeling either the static part (like UML class diagrams [2]), or the dynamic part of a system (like BPMN models [3]). To address the "communication" aspect of the model, both parts need to be taken into account to represent the properties of the real world and their mapping into the IT world. However, the new or changed model should avoid divergences from an available or given domain model that is supposed to serve as a basis for many application models. This is important for the "control" property of a model, to be able to assess the specification, design, and implementation of the future software system. Therefore, compliance with the domain model and realization of specific business logic needs to be harmonized.

In this paper, we propose a new modeling language as part of a method to solve this issue. Our modeling language named BROS (Business Role-Object Specification) can be used to describe and extend a software system that is based on an underlying domain model, and explicitly includes process-driven business logic with respect to its effects on the structural model. In general, BROS is a structural modeling language for design time specification of business objects concerning a (maybe given) domain model as well as specific business logic. Roles fulfilling objects cover the static specification part regarding the separation of concerns, whereas the dynamic specification part of the business logic is expressed via events. The final BROS model serves as a blueprint for development and can be implemented in role-based modeling languages. Although the structural part dominates, we strive for a behavior-aware modeling approach.

The remainder of this paper is structured as follows. Section 2 discusses the related work and current approaches of behavior-aware structural modeling. Section 3 gives an overview of the BROS method, whereas in section 4 we describe the BROS language. Section 5 demonstrates the language with a use case of ordering a pizza, followed by the summary and conclusion in section 6.

2 Related Work

As mentioned before, we strive for a behavior-aware structural modeling method with roles and events. We want to address the issues of using single domain models for various use cases as well as model towards the individual business logic of an enterprise. Especially the tailoring of conceptual business objects of the domain model for different process models is an important feature of modeling software systems.

The role-based paradigm has increased its visibility in the last decades [4, 5]. The static "class" as the main object is not sufficient for modeling business concepts because of its limitation on fixated identity, type, and context. Therefore, various authors propose the role as an appropriate modeling construct in business domains [6–8]. Thus, role-based enterprise modeling is not unknown in modeling.

There are various approaches that model enterprise-specific use cases with roles. In the field of ontologies, many authors state the conceptual foundation of role-based modeling (e.g., [9, 10]). However, conceptual foundations lack implementation details and are only a preliminary stage. In contrast, several works contribute to the methodological aspect of role-based enterprise modeling (e.g., [7, 11]). Kühn et al. present a role meta-model approach of modeling roles for different contexts. However, they do not elaborate in detail on methods of using roles for given business logic.

Colman [12] defines two different ways of representing a role: "player-centric" and "organization-centric", where the latter is focused on the organizational boundary of business logic. Regarding the modeling of information systems, the understanding of "processual roles" is proposed (e.g., in [13]), determining that a role of an object is used as a participant of a certain process. In general, the inclusion of temporal information of business processes is handled in many different ways. Thus, modeling behavior in structural models receives attention in software and system construction. One may use model transformation to solve this issue by transforming process models into structure. However, such approaches do not relate to an underlying domain model as the target structure in general, which is not preferable for template-based system design. A reversed approach includes temporal information into structural models where the modeling of elements like events, situations, and states are in focus. Various papers examine the more abstract, semantic, and conceptual views on events and scenes (e.g., [14, 15]), to increase the foundational understanding of events. However, the details of the method are missing, how one may model an event within a structural model, although that was not the author's intention. Edelweiss et al. [16] propose a method that introduces temporal aspects and object changes to object-oriented models. However, the structural modeling method remains unconsidered. As a relevant related work for BROS, Olivé and Raventós [17] present an approach for defining the concept of an event entity within structural object-oriented models. The authors define the event as a UML stereotype and utilize OCL to formalize post-conditions for events. Nevertheless, by including events into the core model, the reuse of an underlying domain model, as well as modeling towards different business processes, would be difficult. In addition, events as UML stereotypes are still classes, which can be difficult when modeling towards business process models such as BPMN.

3 The BROS Method

Our method was designed to construct and extend software systems by (re)use of domain models. The initial domain model with its business objects may already have been discovered by a domain analysis or given by a domain standard (like in-house templates or reference models). Also, the enterprise-specific business logic should be known. As motivated in the introduction, the BROS method has, in fact, two major core concepts: *(a)* the (re)use of a single domain model to fit different enterprise's business logic, and *(b)* to specify the behavior of objects with respect to the specific business logic. As a result, the final BROS model serves as a blueprint for the construction of an enterprise's software system according to the unique business logic.

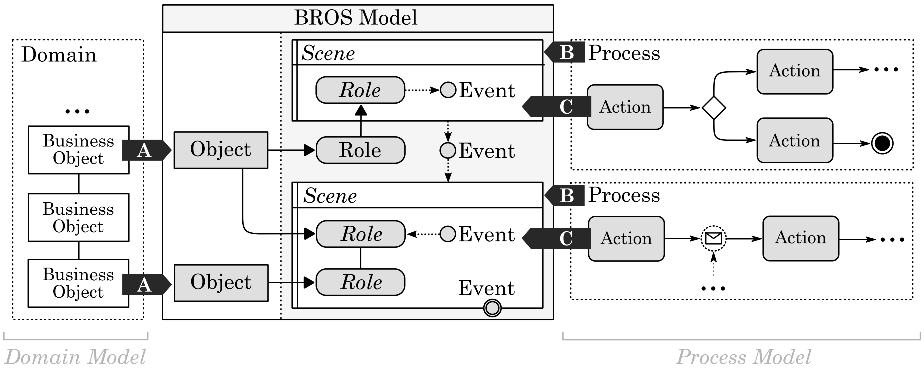


Figure 1. The composition of a BROS model by use of structural and behavioral models

Figure 1 illustrates the BROS method and exemplifies the relationships between the modeling constructs. The objects carry the identity of business objects within the domain ("A" arrows). The (for the enterprise's business logic needed and chosen) objects can be determined by the modeler. Often, the initial domain model is standardized or used by other modelers as well, in which case the modeler must adhere to the domain model as a template, which limits the possibility of introducing new or modifying existent business objects. However, new and changing business logic has to be implemented during the construction and lifetime of the software system. Thus, to express the purpose of (rather static) objects for different (enterprise-specific) business logic in single applications and to support future extensibility, we utilize *roles* as the adaptation construct at design time. A role represents the object and specifies a particular purpose for it. As the objects may be limited to the given domain model, the modeler is free to design roles to fit the objects into the enterprise-specific business logic (e.g., "Meeting Room" as a role of a "Room" object). Even without a limitation to a domain model, roles are useful as a specialization mechanism, since they do not introduce new types and identities for every use case related to the object (e.g., the roles "Meeting Room" and "Dining Room" may use the identity and type of the object "Room" for different use cases). Further, future extensions and evolutions of new functionality are possible by introducing new roles for existent objects.

So far, the specification of objects and roles are the minimum requirements for being a valid BROS model, specifying the static business logic, but not the dynamic behavior. However, for more expressive modeling regarding the enterprise-specific business logic, one can use *scenes* and *events* to describe the behavior and dynamics of roles. A scene is a temporal context for business logic and can often be derived from available process models ("B" arrows). Roles within a scene are acting as participating entities in that process. For example, a "Person in Charge" could be a role within the scene "Room Booking". To specify the lifetime and the participation of roles, we use events as a temporal construct. Events describe a point in time when something has happened (which in turn is based on the enterprise's process model), e.g., "Room Booked". Events are selected by the modeler in such a way as to model the effects of the process on the roles ("C" arrows). The process model that is responsible for the occurrence and impact of events is called the *background process*. In contrast to the domain model, the final BROS model always uses the background process only as guidance, not as a template. That is also why the processes do not need to be BPMN models but may also be of any other behavior model type, e.g., petri nets or sequence diagrams. An extraction of suitable events out of these model types may, however, be different. In conclusion, the BROS model fills a gap between a generic domain model and the specific business logic by providing a possibility to model towards specific requirements of an enterprise's information system.

4 BROS Modeling Language

In this section, we describe the BROS modeling elements in detail. For a formal representation, the complete BROS modeling language is shown in figure 2. We based our meta-model on [18] and modified it towards our requirements.

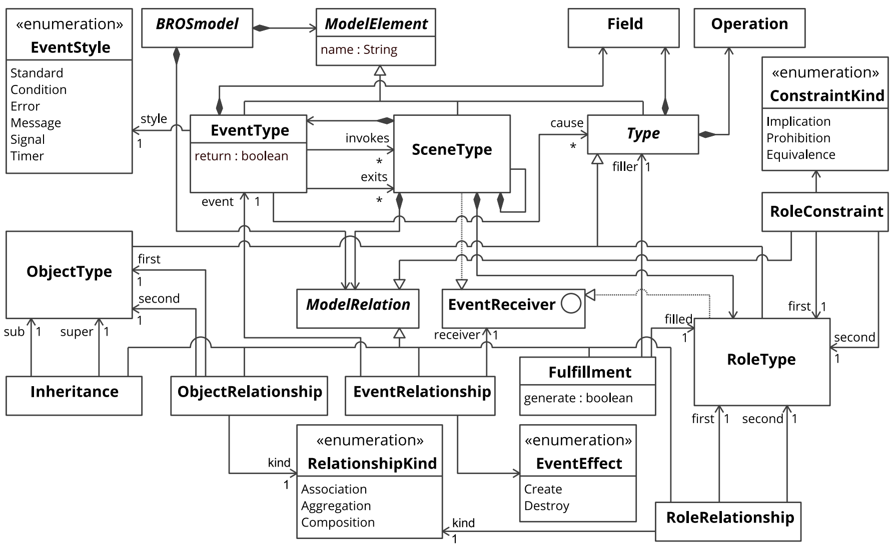


Figure 2. The BROS language meta-model

4.1 Objects

The BROS objects represent the business objects of the model, such as a customer, a rocket, an order, a receipt, or an appointment. The objects are the basic concept for the structural BROS model and are specified using roles, events, and scenes. In theory, we do not need any domain model to start a BROS model. However, if we have such a domain model, we can use its (possibly standardized) business objects as a template for the BROS objects. We define an object as follows:

> *The object is the central modeling construct for representing business objects within a given business domain. The object has an own identity and is uniquely*

identifiable among all other objects. It contains all necessary information to represent the business object with its universally valid properties, behavior, and relationships within the system.

BROS objects are shaped and act similar to the classes from UML. They can be associated with or inherited from other objects and may contain attributes and operations. The reason why we do not call them class but object is simply due to stress that objects always represent a real business object and not any technical class concept. Should the domain model already be in an implementable format (e.g., UML classes with attributes and methods), then we are able to use its classes as BROS objects. However, if this is not the case, one has to create the necessary objects and their relationships in an implementable format.

Please note that the *object* only concerns the object itself. A *compound object* [19], instead, comprises the object as well as all its roles and related events. With the model element of objects, we provide a suitable base for starting the construction of the software system, regardless of whether the object is newly created or already available via an implementable domain model.

4.2 Roles

The definition and understanding of a role in a conceptual model are already well elaborated, e.g., in [20]. In BROS, roles are the static adaptation and configuration parts of business objects. The context may be dynamic (within a scene) or static (the basic BROS model). For both, we use the context definition from Dey [21]:

"Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." [21]

We interpret this definition in such a way, that a scene (and the model itself) is a particular "information" (the "context") that characterizes the situation of objects (the "entities") whereas the objects then are relevant for the general interaction within the scene, expressed as roles as context participants. Our definition of a BROS role is based on the context definition:

A role is a contextual modeling construct with state and behavior that is fulfilled by an object or its roles to represent it in the user's context and extend or change its corresponding specifications and interactions.

A role can be fulfilled either directly by the object or indirectly by a role that is already fulfilled by the object. In our definition, at the type level, the role is "fulfilled by" an object as a *player*. However, at the instance level, the object instance "plays" the role instance (respectively the role is "played by" its player). A fulfilled role shares the related player's identity at runtime. Graphically, a role is represented by a rectangle with curved edges. Figure 3 shows a role fulfillment.

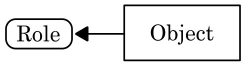


Figure 3. A player object fulfills a role

Role Relationships. In order to successfully represent business logic and unique structure, different relationship kinds are needed to represent the interaction between two elements. Roles can be related to each other to represent their related business logic and interactions, based on the UML language. This includes the kinds shown in table 1.

Further, BROS also supports the usage of role constraints, introduced by [5]. We may use the constraints *Implication*, *Prohibition*, and *Equivalence* to specify constraints between two roles.

Table 1. Role relationship kinds

Name	*Icon*	*Description*
Association		interaction-based connection
Composition		"contains" connection
Aggregation		"owns" connection (special case of composition)
Implication		role A implies playing role B
Prohibition		role A and B cannot be played at the same time
Equivalence		role A and B can only be played together

Role-based Adaptation. To redefine several aspects of an object in a specific context, roles can be fulfilled to alter its structure or behavior. For that, we use the delta concept, based on the ideas of delta programming [22]. We only use the *add* (+) and *delete* (–) actions, as those two are logically sufficient to represent all other modifications. For convenience, we also allow *modify* (~) as a combination of both. A role, therefore, may use add, delete or modify as *adaptation actions* to alter properties of its player object. The things we may redefine with roles are manifold: attributes, visibility, methods, parameter, types, or other type-atomic elements. However, we never alter the actual object, but only its roles hold the delta information. In fact, with roles, it is possible to represent (and modify) the original object aligned to the specific requirements of the business logic. Nevertheless, it is not mandatory to specify any delta for using roles. A role may be empty in its delta, but can still be used to express certain business logic, e.g., a state of an object. Note that a given delta implies knowledge about the (future) player and its properties to which the delta refers. If new players are added, the delta may require changes. Figure 4 shows an example of an adaptation.

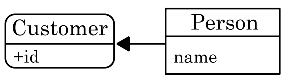


Figure 4. A role changes the player's properties (adds "id" as delta)

4.3 Events

We use events as an essential mechanism for the dynamic aspect of roles. In the UML, events are introduced as “something that may occur at a specific instant in time. One event may have many occurrences, which may happen at different times.” [2] We share this definition and extend it further:

> *An event is a modeling construct for specifying the temporal behavior of roles. It occurs at a specific instant of time, with possible reoccurrences. An event may have causes to define the objects that may trigger the event and determines the object fulfillment of roles.*

Event modeling is a frequent topic in conceptual modeling, mainly in process models [23]. However, in BROS, we explicitly focus on the impact of process events on the objects (and their roles) of structural role-based models. Such events are, as described in section 3, interfaces to business process models and represent the impact of process models in the structural model. Events are context-dependent, which means that they only apply in the context and can be triggered by objects or roles in the same scene in which they are located.

When modeling an event, it appears as a simple circle within the diagram. At design time, an event in the model means that it can occur in the course of business logic and certain effects regarding roles result from it. However, at runtime, an event is something that must be triggered, for example, by a method call. As a result of events, roles are created and dropped by the *event effect*. Each role created and orchestrated by an event is called a *dynamic role*, whereas roles that occur independently of events in the system are called *static roles*.

Event Styles. Events are also known from BPMN and described as “something that happens during the course of a process”. [3] The BPMN specification differentiates several types of events like message, error, or cancel. As such a classification of events is useful for business modeling, BROS uses some of them within its specification, too: standard, message, timer, error, condition, and signal. Those *event styles* are shortly represented in table 2. However, BROS uses slightly different semantics. Whereas the BPMN event types are focused on the different participant actions, in BROS an event only types the origin of something that happens (e.g., in BPMN, the timer event is involved in an actual workflow, and in BROS it is triggered by an “invisible clock”).

Table 2. General event styles

Name	*Icon*	*Description*	*Example (Pizza Ordering)*
Standard		- standard event	customer added
Message		get message event	order received
Timer		get timer event	pizza 20 minutes baked
Condition		get condition event	pizza less than 40°C

Error		get unexpected event	no more salami available
Signal		get signal event	successful credit card check

Event Effects. An event defines a specific instant in time at which a role instance is created or destroyed. This enables orchestration of the roles at runtime. If an event is triggered, it handles the roles as a consequence (the event's effect). The event's effect options are shown in table 3. Although events can only be triggered within their context, event effects can in principle have roles outside of their own context as a target.

Table 3. Event effects

Name	*Icon*	*Description*
Create		bind a new role to an object
Destroy		unbind a role from an object

4.4 Scenes

The grouping of roles that belong to a specific (temporal) task is called a scene, adapted from [15, 24]. A scene is the main element for describing the execution of any business logic that affects the roles. We define a BROS scene as follows:

> *A scene is an instantiable temporal collaboration context of roles and events, related to the same business logic part.*

Good indications of such scenes are activities or operations, e.g., "student enrollment" or "order pizza". The scene does not model the business process of such tasks but only the role composition and lifetimes. When defining roles for a scene, it is important to always focus on the scene and its required roles as participants, not the existing business objects and the roles they might be able to use. Thus, the background process acts as a hint but does not give a complete answer on the needed roles for scenes. Further, a scene may vary in its granularity. A rather coarse-grained scene would be "book flight" while a fine-grained scene would be "pay with credit card," possibly also as a sub-scene of the former one. However, events are not limited to scenes. Due to readability, the fulfillment arrows do not cross the edge of a scene, but note the target role name, as shown in figure 5. If an object instance is newly created as soon as it takes part in a certain scene, it is marked with a double arrowhead, also shown in figure 5. That arrow is called a *generate fulfillment* and only possible for fulfillments targeting dynamic roles.

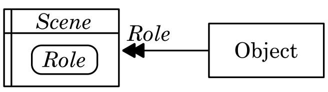
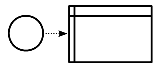
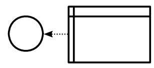
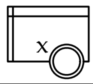


Figure 5. A new object instance is created as soon as the role is created in the scene

Special Event Functions. Scenes also define some other special functions for events and roles to better integrate the business logic flow, which is summarized in table 4.

Table 4. Scene event functions

Event Function	*Description*	*Notation*
Invoker Event	specifies the start of a scene (invokes the scene); all *init roles* of the scene are created	
Exit Event	specifies the termination of a scene from the outside; the scene terminates, and its roles are destroyed; invoked scenes are interrupted (exit)	
Return Event	specifies the end of the scene from the inside; all roles in the scene are destroyed	

A scene may start by any possible *invoker event*. At runtime, multiple instances of a scene, started by possibly different invoker events, may exist. *Init roles*, which occur with the invoker event, are particular roles within a scene that get played (like a parameter) as soon as the scene is invoked. There is always a multiplicity given with a role to determine the number of possible instances for that role at runtime. An init role has a lower bound of at least one instance (e.g., “1..*”).

Exit events are events, which are triggered outside of the scene. A triggered exit event is like an external interrupt. All roles within the scene are destroyed, and the scene is over. However, the actual exit event may have some further effects on other roles as well as causes.

The standard way of exiting a scene is using the scene’s *return events*, a double lined circle on the scene’s edge. They are for the intended outputs of the business logic, even errors or unexpected behavior. The difference between return and exit events is that return events are always triggered within the scene and from the business logic itself, not from the outside. A return event’s effect will always take place, regardless of the invoking context. The background process may indicate one or more return events, but it could also be determined by the modeler due to technical reasons.

5 Modeling Case Study

In order to demonstrate our approach, we present a business use case regarding different mechanisms of BROS. As a foundational background process, we use a very simplified “order pizza” BPMN example, as shown in figure 6. It describes a simple pizza ordering within a restaurant plus a routine for customer feedback.

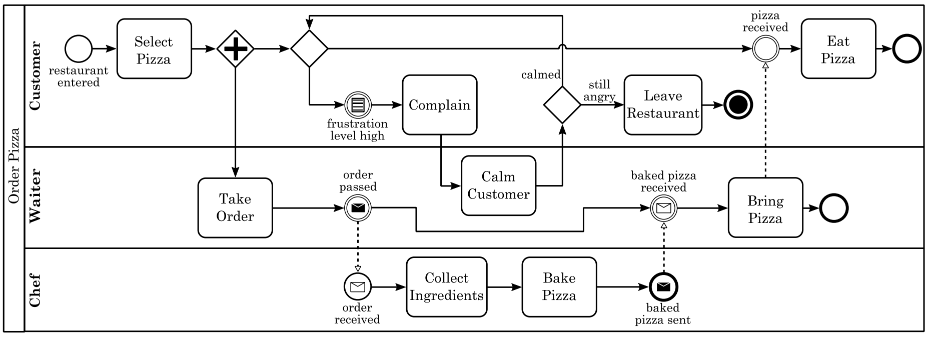


Figure 6. The background process for this case study

It is not possible to translate a BPMN directly into a BROS model. Instead, the construction of BROS models uses BPMN as the background process. We rather assume that a basic structural domain model already exists and is adapted to fit a specific use case or process. Figure 7 shows the final BROS model. We omit attributes and operations due to better readability.

<BROS Example Restaurant>

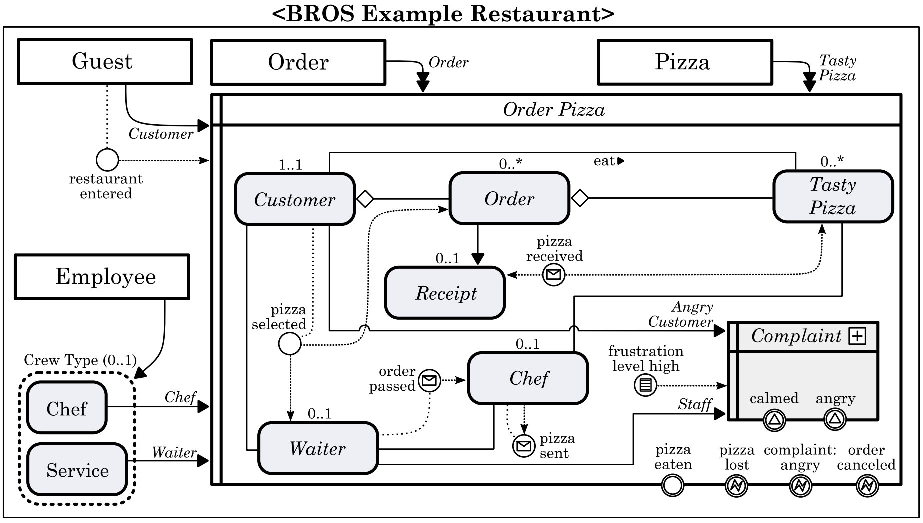


Figure 7. The “Order Pizza” modeled with BROS

Context and Objects. For this example, we use static contexts (the named BROS model), dynamic contexts (scenes), and the corresponding objects out of the business object set given within the domain model. This is either done by the modeler’s requirements or by using objects provided by a standardization, e.g., a reference model for the specific domain. We identified the context of the *Order Pizza* model first:

- The background process example describes the pizza ordering process. Thus, our domain is the restaurant, modeled as the *Example Restaurant* BROS model.

- We used four objects: *Employee*, *Guest*, *Order*, and *Pizza*. They serve as the most founded elements and carriers of identities. We assume them to be given within our domain model.
- Our background process implies a scene called *Order Pizza* (probably only one of many scenes within a restaurant). We decided to implement another sub-scene *Complaint* (instead of a single event) due to any existing complexity in the background process related to this.

The additional scene is *folded*, as we are not interested in its inner role behavior but only the players and the outcome. However, other models still might use the full representation of these scenes.

Roles. The roles adapt the objects, tailored for our own (enterprise-specific) business logic of ordering a pizza. Often, the roles can be derived from the given participants of the related background process (e.g., a swim lane in BPMN), which interact with each other in that specific scene. However, one should not forget to also define roles for things and not only for persons.

- We refined the *Employee* object with two static roles: *Service* and *Chef*. Both roles are not part of our scene (means independent of ordering a pizza). We grouped them as *Crew Type*. Every *Employee* has to play at most one role of the *Crew Type*, marked by the multiplicity.
- The *Order Pizza* scene has the following dynamic roles to fulfill its desired business logic: *Customer*, *Waiter*, *Order*, *Chef,* and *Tasty Pizza*. The *Customer* is an init role.
- One needs to define players for the scene's roles. Thus, we use given objects as players. Further, *Order* and *Pizza* use the generate fulfillment so that it is clear that these generate a new instance when taking on the role within the scene.
- Since *Complaint* is folded, we imply knowledge about the inside roles. Thus, a *Waiter* fulfills the role of *Staff* and *Customer* fulfills the *Angry Customer*.
- There is a dynamic role called *Receipt* to respond to the finished order in the form of "pizza received". This role will be played indirectly by the *Order* object.

Please note how the role *Chef* is fulfilling the role of *Chef* again. This is due to the different context. The outside *Chef* has, logically, nothing to do with the *Chef* within the scene. The former role states a crew member, the latter states a person who bakes the pizza. This may be different in other scenes involving the crew's *Chef*. At runtime, both *Chef* roles use the same identity as the related *Employee* instance.

Events. The adding of events depends on the modeler's intention and the background process and can differ widely. We chose one possibility for our use case.

- We stated *restaurant entered* as the invoker event for the *Order Pizza* scene.
- The start and end of a participant on a swim lane in BPMN could provide information about BROS events that start and end a role. For example, we used *order passed* as the start event for the *Chef* role, which ends with *pizza sent*.

- We modeled several return events that end the *Order Pizza* scene. Further, also the returns of the folded scene *Complaint* were defined.
- We used the *Receipt* role as the event effect of *pizza received* to generate the receipt for the customer.
- Finally, optional event causes are added. For example, we modeled that any *pizza sent* event may only be triggered by *Chef*.

Remarks. In general, we used the process, described with BPMN, as a reference for the final structural model, described with BROS. Our BROS implementation is only one possible option. The extent of modeling (and thus the richness of detail and readability) depends on the modeler's requirements and needed scope and granularity. Although the final model may be more complex and larger, it is a detailed adaptation of a (possibly already existing) structural model for a specific business logic use case. Additionally, for example, we could add an "order pizza online" or "payment" scene within the same domain without changing the underlying domain model. Please note that in order to make the model implementable, the (delta) attributes and operations would now have to be added.

6 Conclusion

In the context of role-based modeling, event modeling, and temporal behavior, we have proposed BROS, a new modeling language for behavior-aware structural modeling. It utilizes objects, roles, events, and scenes to describe business objects as role objects, orchestrated by events to represent business logic.

The BROS model reconciles the disparity between a conceptual domain model and individual business logic of the target context. Thus, the BROS method comprises three steps: *(a)* refining of the conceptual model for implementation-ready objects, *(b)* using roles as the variable part of objects for tailoring them towards individual business logic, and *(c)* defining a temporal context regarding business logic in the form of process models, guiding the occurrence of roles and their interaction. We defined a language meta-model and used a fictional case study to show the application of our BROS method in principle and in order to demonstrate feasibility. We consider that the best application of BROS occurs at an early stage of software development when the domain model's expressiveness turns out to be a limitation and software engineers are in need of further details regarding the targeted business logic that has to be fulfilled. On the basis of this, we argue that the possibility of detailed and expressive modeling is helpful for defining new and changing functionality of a software system. Nevertheless, this is also a limitation of BROS, since a BROS model of a comprehensive software system would be too complex. Thus, BROS is possibly better suited for modeling only an excerpt of the complete system, as BROS roles and scenes allow the separation of concerns. However, it is not appropriate to provide a complete system overview. Therefore, a combination of BROS (for detailed views) with a previous model instance (e.g., the overall domain model) works best.

However, some future work is still needed. BROS is currently in its first version. Some modeling scenarios might not yet have been considered, or there might exist some inconsistent features. Further, BROS is formalized by a meta-model, but a concrete syntax (e.g., notation) is also needed. We want to enhance the event element so that more functionality of process models can be covered. The integration of pre- and post-conditions of events also has its benefits [17]. Eventually, our approach needs to be tested in contexts closer to real applications. We also intend to do a broader evaluation and study of the resulting BROS models regarding their applicability and usability (e.g., with [25]) in practice. Additionally, we want to develop tool support and concrete guidelines on how to create BROS models concerning existing background processes.

7 Acknowledgements

This work is funded by the German Research Foundation (DFG) within the Research Training Group „Role-based Software Infrastructures for continuous-context-sensitive Systems" (GRK 1907).

References

1. Furrer, F.J.: Future Proof Software Systems. Software Technology Lecture Talk. Dresden, Germany (2017).
2. Object Management Group: Unified Modeling Language v2.5.1. (2017).
3. Object Management Group: Business Process Model and Notation v2.0. (2011).
4. Steimann, F.: On the representation of roles in object-oriented and conceptual modelling. Data Knowl. Eng. 35, 83–106 (2000).
5. Riehle, D., Gross, T.: Role Model Based Framework Design and Integration. In: Proceedings of the 13th ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications, OOPLSA. pp. 117–133. ACM, New York, NY, USA (1998).
6. Cesare, S. de, Henderson-Sellers, B., Partridge, C., Lycett, M.: Improving Model Quality Through Foundational Ontologies: Two Contrasting Approaches to the Representation of Roles. In: Jeusfeld, M.A. and Karlapalem, K. (eds.) Advances in Conceptual Modeling. pp. 304–314. Springer Cham, Stockholm, Sweden (2015).
7. Zhao, L.: Designing Application Domain Models with Roles. In: Assmann, U., Akşit, M., and Rensink, A. (eds.) Model Driven Architecture. pp. 1–16. Springer Berlin Heidelberg, Twente, The Netherlands (2005).
8. Frank, U.: Delegation: An important concept for the appropriate design of object models. J. Object Oriented Program. 13, 13–17 (2000).
9. Almeida, J.P.A., Guizzardi, G., Santos, P.S.J.: Applying and extending a semantic foundation for role-related concepts in enterprise modelling. In: Proceedings of the 12th IEEE International Enterprise Distributed Object Computing Conference, EDOC. pp. 31–40. IEEE (2009).
10. Steimann, F.: Role = Interface : A Merger of Concepts. J. Object-Orientated Program. 23–32 (2001).
11. Zhao, L., Kendall, E.: Role modelling for component design. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. IEEE, Maui, HI, USA (2000).
12. Colman, A.: Role Oriented Adaptive Design, (2006).

13. Boella, G., Steimann, F.: Roles and Relationships in Object-Oriented Programming, Multiagent Systems and Ontologies. In: Cebulla, M. (ed.) Object-Oriented Technology - ECOOP 2007 Workshop Reader. pp. 108–122. Springer Berlin Heidelberg, Berlin, Germany (2007).
14. Galton, A.: States, Processes and Events, and the Ontology of Causal Relations. College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK (2012).
15. Almeida, J.P.A., Costa, P.D., Guizzardi, G.: Towards an Ontology of Scenes and Situations. In: 2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). pp. 29–35 (2018).
16. Edelweiss, N., Oliveira, J.P.M. de, Pernici, B.: An object-oriented temporal model. In: Rolland, C., Bodart, F., and Cauvet, C. (eds.) Advanced Information Systems Engineering. pp. 397–415. Springer Berlin Heidelberg, Paris, France (1993).
17. Olivé, A., Raventós, R.: Modeling events as entities in object-oriented conceptual modeling languages. Data & Knowledge Engineering. 58, 243–262 (2006).
18. Kühn, T., Leuthäuser, M., Götz, S.: A Metamodel Family for Role-Based Modeling and Programming Languages. In: Combemale, B., Pearce, D.J., Barais, O., and Vinju, J.J. (eds.) Software Language Engineering. pp. 141–160. Springer, Västerås, Sweden (2014).
19. Leuthäuser, M.: A Pure Embedding of Roles, http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-227624, (2017).
20. Bera, P., Burton-Jones, A., Wand, Y.: Improving the representation of roles in conceptual modeling: theory, method, and evidence. Requirements Eng. (2017).
21. Dey, A.K.: Understanding and Using Context. Pers. Ubiquitous Comput. 5, 4–7 (2001).
22. Damiani, F., Schaefer, I.: Dynamic delta-oriented programming. In: Proceedings of the 15th International Software Product Line Conference, SPLC. ACM, New York, New York, USA (2011).
23. Barros, A., Decker, G., Grosskopf, A.: Complex Events in Business Processes. In: Abramowicz, W. (ed.) Business Information Systems. pp. 29–40. Springer Berlin Heidelberg, Poznan, Poland (2007).
24. Guarino, N., Guizzardi, G.: Relationships and Events: Towards a General Theory of Reification and Truthmaking. In: Adorni, G., Cagnoni, S., Gori, M., and Maratea, M. (eds.) AI*IA 2016 Advances in Artificial Intelligence. pp. 237–249. Springer Cham, Genova, Italy (2016).
25. Krogstie, J.: SEQUAL as a Framework for Understanding and Assessing Quality of Models and Modeling Languages. Int. J. Inf. Syst. Model. Des. 3, 24–45 (2012).

Generating Smart Glasses-based Information Systems with BPMN4SGA: A BPMN Extension for Smart Glasses Applications

Jannis Vogel[1] and Oliver Thomas[1]

[1] University of Osnabrueck, Chair of Information Management and Information Systems, Katharinenstr. 3, 49074 Osnabrueck, Germany
{jannis.vogel,oliver.thomas}@uni-osnabrueck.de

Abstract. Although smart glasses allow hands-free interaction with information systems and can enhance business processes, they face problems with the adoption in businesses. Implementation challenges arise due to specific hardware conditions e.g. computational power, limited battery, small screen size and privacy issues caused by the camera. In addition, not many programmers are specialized for the development of smart glasses-based applications to conquer the mentioned challenges. We address this issue with a generation tool for smart glasses-based information systems. A BPMN extension for smart glasses applications allows the abstract specification. Specified processes are then integrated into a model-driven software development approach that transforms processes directly into smart glasses applications. This paper covers the design and development phase of the abstract and concrete syntax of the BPMN extension and the representation of the architecture to generate smart glasses-based information systems with the new developed BPMN extension.

Keywords: Smart Glasses, BPMN Extension, Model-driven Software Development, Domain-specific Modelling Language.

1 Introduction and Motivation

Due to the mobility, hands-free interaction with information systems and a range of different functionalities, smart glasses are suitable for various application areas [1]. Widespread applications are pick-by-vision [2], remote assistance [3] and step-by-step guidance [4]. In particular smart glasses are suitable as a service support system to guide users through a complex working process such as in the technical customer service domain [4]. In general, smart glasses are applied in businesses to enhance business processes due to information provided in the user's field of view.

However, different obstacles arise during the implementation of smart glasses in businesses. Hobert and Schumann identified 25 challenges with the integration of wearables (i.e. smart glasses and smart watches) into businesses through 21 semi-structured interviews. They recognized that organizations have an urgent demand for software developers that know the implementation of wearables into businesses.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Furthermore, interviewed experts in the study mentioned that only a few wearable third-party experts exist [5]. This circumstance hinders the adoption of smart glasses and hampers an adequate dealing with smart glasses specific implementation barriers such as the small screen size [6], limited battery capacity and privacy issues [5].

A promising approach to overcome the mentioned obstacles is the usage of domain-specific modelling languages (DSML) to design software [7], in this case smart glasses applications, and to steer the implementation process with a model-driven software development (MDSD) approach [8], i.e. models have a direct effect on the desired software. During the research project Glasshouse [9] we followed that approach by using the modelling language BPMN to describe smart glasses-based use cases in the logistics domain. Thereby, we identified that existing process modelling languages such as EPC or BPMN are not suitable for the technological representation of smart glasses functionalities [10]. Consequently, a proper model-based design of smart glasses applications and their utilization in a MDSD approach was not feasible.

Hence, we address implementation obstacles and the non-technical expressiveness of process modelling languages in terms of smart glasses with the development of an abstract and concrete syntax of a BPMN extension for smart glasses-based information systems. Thereby, adequately represented smart glasses-based process models lead to smart glasses-based information systems by a MDSD concept. The BPMN extension supports a precise technological representation of smart glasses functionalities into business processes and the MDSD approach allows non-technical experts the design and readjustment of smart glasses-based information systems. Thus, the superordinate research question is: *How can smart glasses-based functionalities be integrated within a BPMN extension, whose models allowing the generation of smart glasses applications with a model-driven software development approach?*

The paper is structured as follows: In the first section, we declare the relevance, motivate for a BPMN extension and explain the connection with a generating system for smart glasses-based information systems. In the second section, related work regarding smart glasses applications in the domain of process modelling is explained. Further, the objectives of the solution are outlined. Afterwards we describe the applied research method in section three. Excerpts from the results of the applied research method are presented in section four. In the next section, the implemented architecture for the MDSD approach is explained. Section six mentions discussions and limitations points that appeared during the research project and sums up the article.

2 Smart Glasses Applications related to Process Modelling and Objectives of BPMN4SGA

In the range of smart glasses applications, some have a close connection with process modelling in various forms. For instance, Metzger et al. [11] developed a smart glasses-based modelling system to model service processes during process execution. The worker uses specific speech commands and can take pictures with the smart glasses camera to create a process model. In addition, modelled processes can be used as step-by-step process guidance. Furthermore, Fellmann et al. [12] propose a process modelling recommender system to accelerate process modelling activities with smart glasses. A different approach presents Petersen et al. [13], their application allows the real-time elicitation of work processes with videos from the smart glasses user's view. The videos being used to generate an augmented reality manual that presents correct performed tasks with a green coloring of the worker's hands.

In comparison with the mentioned smart glasses applications, BPMN4SGA with an additional generating system provides step-by-step guidance based on process models like in [11], too. Though a distinction is the abstraction level, the smart glasses-based modelling system by Metzger et al. [11] uses simple workflow patterns like sequences or exclusive choices with the dedicated activities, whereas BPMN4SGA's objective is to support smart glasses functionalities within smart glasses information systems, e.g. to use the smart glasses camera or specific voice commands. Generally, BPMN4SGA allows a more detailed description of smart glasses applications. Besides a more precise description, other advantages come with the generation system and the applied MDSD approach.

This leads to several use cases that result in specific objectives of the solution. One use case is in the area where companies with no IT departments facing difficulties to adopt smart glasses from a technical point of view due to dependencies from third-party developers [5]. This circumstance ensures a rejection of new technologies such as smart glasses, although the advantages of the technology are well-known. A generating system for smart glasses-based systems can encounter technological rejection that is based on difficulties in software development. Therefore, especially small and medium-sized companies should benefit from the solution. Furthermore, smart glasses in a digitalized and mobile environment needing a fast readjustment of their information systems that can be accomplished with the proposed solution. Another tackled aspect is the reduction of investment costs and -risks into smart glasses-based systems. Assuming that one company provides the solution, economies of scale could be achieved that lead to lower and predictable costs. Particularly intangible benefits of smart glasses systems are hard to collect that makes a clear investment decision challenging [14]. BPMN4SGA and the associated generating system allow testable and fast running smart glasses systems that help to identify intangible benefits beforehand.

3 Research Method

The overall research project applies the Design Science Research Method (DSRM) by Peffers et al. [15]. This paper partly covers the design and development phase. For the actual phase, we use the method for the domain-oriented development of BPMN extensions by Braun and Schlieter [16] that extends the method for the technical development of BPMN extensions by Stroppi et al. [17] that is shown in figure 1. The domain-oriented analysis is covered in [10]. The outcome was the identification of 19 extension requirements for a BPMN extension that allows the representation of smart glasses-based processes, which can be transformed into smart glasses-based information systems with a MDSD approach.

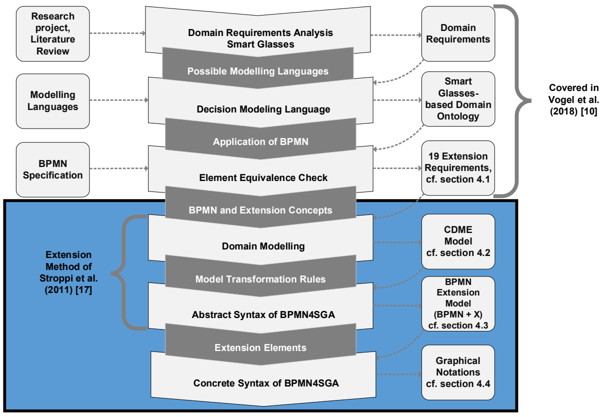


Figure 1. Applied method for the domain-oriented development of BPMN4SGA in accordance with Braun and Schlieter [16]

This paper focuses on the following steps highlighted in blue, which are overall three steps. The first step covers a domain model which is represented by a Conceptual Domain Model of the Extension (CDME) that is depicted by a UML class diagram. The CDME is the groundwork for the abstract syntax definition of the BPMN extension that is represented by a BPMN extension model (BPMN+X) and is also depicted by a UML class diagram. For the conversion of the CDME into the BPMN+X special model transformation rules are applied. Both steps are covered by the extension method of Stroppi et al. and are explained in detail in [17]. After that, the concrete graphical syntax is developed based on the BPMN+X diagram. Further design recommendations have to be covered. For instance, the shape of standard BPMN notation elements should not be altered [16].

4 Design and Development of BPMN4SGA

4.1 Extension Requirements for the BPMN Extension

Based upon insights from a one-year project phase for the identification of potential use cases in the logistic domain [9] and design guidelines from literature regarding smart glasses-based systems, we listed in an in-depth analysis technical- and software-based requirements for a dedicated BPMN extension that allows representing smart glasses-based information systems [10]. Thereby, we identified six technical-based (TB-DR) and six software-based (SB-DR) domain requirements (DR) (cf. table 1). In order to link the derived requirements with BPMN concepts, a further developed smart glasses-based domain ontology represents BPMN concepts and their linking with the developed DR. Apart from its use as a representation of knowledge, the ontology was used to steer the element equivalence check. In doing so, DR were compared with directly linked BPMN concepts. Thereby, the element equivalence check revealed 19 extension requirements where 14 requirements were not equal (NE) and 5 were conditional equail

(CE) regarding compared BPMN concepts. For all requirements occurs a specific extension, a graphical representation as an extension format is designated for smart glasses activities and their event types. Further, additional domain concepts that are not covered by standard BPMN are extended through new attributes. The design decisions were made based on the level of equality of domain to BPMN concepts and their novelty. This led only to new graphical notation elements for events and activities that will simplify learning BPMN4SGA.

Table 1. Element Equivalence Check and Extension Format for BPMN4SGA [10]

DR	Domain Concept	BPMN Concept	Equal?	Extension Format
TB-DR1	Activity and event support by the tracking concept	Activity Event	NE NE	Graphical Graphical
TB-DR2	Interactions steer the process flow, can trigger events and confirm activities.	Process flow Event Activity	NE NE NE	Attribute Graphical Graphical
TB-DR2	Manual decision through interaction inputs	XOR Gateway	CE	Attribute
TB-DR3	Visualisation of the process, brightness, and display size	Process flow Notation elements	CE	Attribute
TB-DR4	Camera to take pictures or videos	Activity	NE	Graphical
TB-DR5	Context via sensors	Event	NE	Graphical
TB-DR6	Communication	Activity Event	CE CE	Graphical Graphical
SB-DR1	Distractive visualisations	Process Flow	NE	Attribute
SB-DR2	Information provision Content type	Activity Activity	NE NE	Graphical Attribute
SB-DR3	Process improvements	Process	NE	Attribute
SB-DR4	Additional information	Data Object	CE	Attribute
SB-DR5	Object identification	Activity Event	NE NE	Graphical Graphical
SB-DR6	Communication system	Activity	NE	Attribute

4.2 Conceptual Domain Model of the Extension (CDME)

The Conceptual Domain Model of the Extension (CDME) represents the domain and BPMN concepts and their relationships (cf. figure 2). The CDME includes BPMN concepts and extension concepts that are derived from the identification of extension requirements (cf. section 4.1). Thereby, BPMN concepts are also covered in the BPMN metamodel and extension concepts, filled in grey, include concepts from the smart glasses domain [17]. The connection between standard BPMN concepts and extension concepts occurs through generalization relationships. The new graphical extension concepts have a generalization relationship to existing BPMN concepts e.g. TrackingTask has a generalization relationship to the BPMN concept task and are the new introduced notation elements (cf. section 4.4). The generalization relationships between BPMN concepts to extension concepts like GraphicalRepresentation, ProcessImprovement, SteeringProcessSetting, VisualisationSetting, ManualDecision and AdditionalInformation should add additional attributes to existing BPMN concepts. For instance, GraphicalRepresentation class attributes should be included in any task.

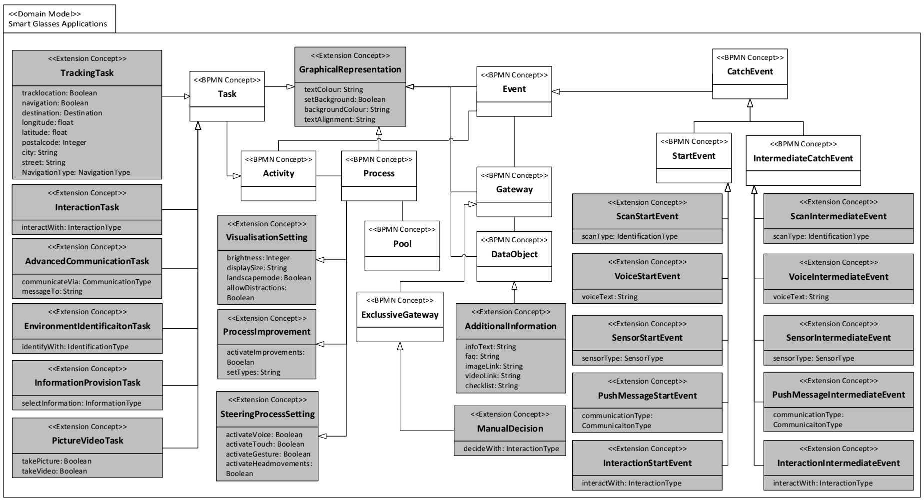


Figure 2. Conceptual Domain Model of BPMN4SGA

In summary, the CDME consists of six new tasks that have further unique attributes. They are getting complemented with distinctive graphical attributes such as text colour, background colour or alignment of the text content. Furthermore, additional information can be attached at tasks through data objects to provide the user with additional information if they request it. Events can be either catch or throw events. Thereby, catch events catching a trigger, whereas throw events throwing a result that can be caught [18]. We only covered catch events for several reasons. Firstly, the user throws events indirectly through the smart glasses functionalities that have to be caught by start or intermediate events. When the worker uses smart glasses in the work environment – he triggers different kinds of events that are also indirectly covered in the smart glasses specific tasks. Encountering this problem, we decided to avoid a doubling and we do not cover throw events in the BPMN extension. Further, with this division, a clearer understanding of the process flow is realized. Start events of the extension model can catch a trigger and starting a smart glasses-based process. This allows a technical selection of processes through e.g. identification of objects through scans, voice-based commands, sensors, interaction inputs or push messages.

Intermediate events can interrupt the process until a specific event is thrown. Additionally, intermediate events can steer the process flow when an event-based gateway is used. On the one hand, we are aware that the decision not to include throw events is debatable and have to be observed during the evaluation phase. On the other hand, it allows no confusion regarding the differences between smart glasses-based tasks and thrown events. Further comparable BPMN extensions, like uBPMN that represents ubiquitous computing processes, do only cover catch events too [19]. The manual decision class represents an exclusive decision that can be handled by the smart glasses user by interactions such as voice, touch or gestures. Further, the different kinds of process steering settings can be enabled in the process settings. Finally, process improvements can be activated to enable the functionality to retrieve feedback from the smart glasses user regarding the work process, and visual and graphical settings set the general design of the information system. Thereby, the CDME includes all listed requirements (cf. table 1) and brings the concepts in a sufficient order.

4.3 BPMN Extension Model (BPMN+X)

The evolved CDME is transformed into the BPMN extension model (BPMN+X) (cf. figure 3). BPMN+X is a specific language developed by Stroppi et al. [17] that is based on the BPMN extension mechanism [18]. Classes covered by the BPMN metamodel are in white and classes representing the extension domain are in grey. To evolve the BPMN+X model, we applied the transformation rules defined by Stroppi et al. [17]. Especially rules five and seven are used to represent a CDME generalization relationship in the BPMN+X model. Rule 5 implied that the BPMN generalization relationship remains by two BPMN concepts. In contrast, if a generalization relationship exists in the CDME with an extension concept, then the relationship is presented in the BPMN+X model through an ExtensionRelationship. According to Stroppi et al. [17], an ExtensionRelationship represents the conceptualization of extensions and their connected extensions are understood as customization of existing elements from the BPMN metamodel [17]. Further extension enumerations (Extension Enum) are represented for new attribute types e.g. SensorType.

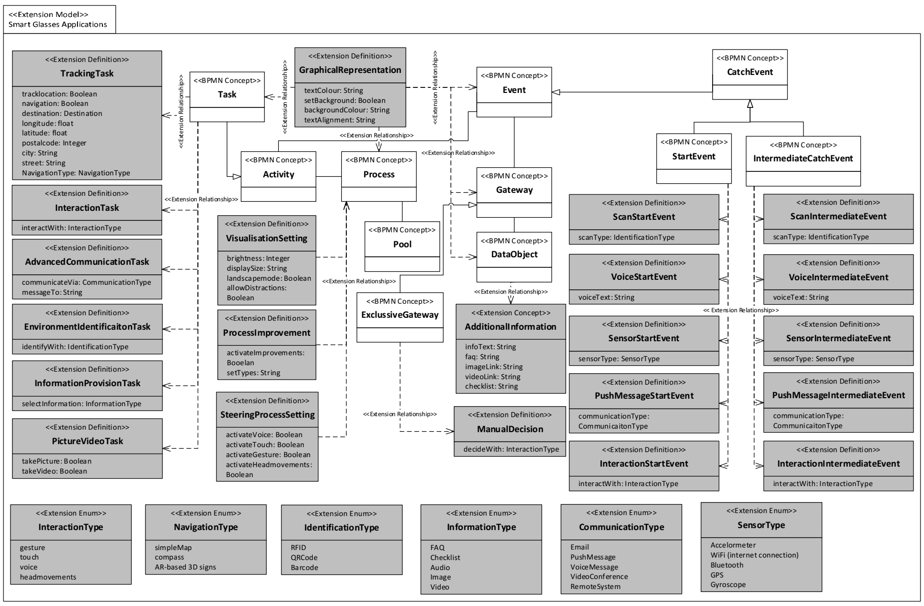


Figure 3. BPMN Extension Model (BPMN+X) of BPMN4SGA

In a next step, Stroppi et al. [17] suggest two further transformation steps. The created BPMN+X can be used to create an XML schema extension definition model that can be transferred into an XML schema document. Similar to the development in [20], we do not execute these steps, because the research project focuses primarily on the development of a DSML. To cover the extension model into a new visual BPMN extension, we propose a graphical representation of the new tasks and events (cf. section 4.4).

4.4 Graphical Representation of BPMN4SGA

The graphical representation of BPMN4SGA is represented through additional events (cf. figure 4) and activities (cf. figure 5). Further extension elements are represented through additional attributes. We include the design guidelines for DSML by Frank [21] to design the specific graphical notation elements. Because of the usage of standard BPMN shapes and just the alteration of symbols, we indirectly fulfilled all design guidelines, except design recommendations in regard to symbols. Therefore, we checked popular icon databases such as NounProject [22] to identify common icon designs for specific word phrases. Assuming that many persons use these icons to represent specific domains an approximation will increase the understandability of the graphical notation of BPMN4SGA. Based upon a selection, we adapted the icons for smart glasses applications and created scalable vector graphics (SVG).

In summary, ten start and intermediate events are added as graphical elements. Figure 4 shows the new events of the BPMN extension for smart glasses applications. The sensor (1), voice (2) and scan (3) start and intermediate events are in accordance with the uBPMN [19]. Sensor-based events (1) react to build in smart glasses sensors. For instance, the Vuzix M300 is equipped with the sensors: acceleration, Bluetooth, internet connection, GPS or gyroscope [23]. Voice-based events (2) capture voice commands and can start new smart glasses processes and can have an impact on the process flow. The same applies to scan-based events (3) regarding the identification of objects e.g. through QR- or barcodes. Start and intermediate events catch specific user-generated interaction-based events (4). These could be gestures, voices, touches on buttons or head movements. Voices are also covered by this extension element to depict combinations with other user-generated interactions. Push message-related events (5) trigger new smart glasses processes for a user or can have an impact on the process flow by using communication systems. They allow in combination with the advanced communication task the collaboration between other smart glasses users.

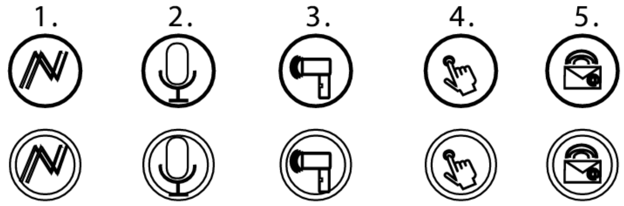


Figure 4. Events of the BPMN Extension for Smart Glasses Applications (BPMN4SGA)

The smart glasses specific tasks are depicted in figure 5. The tracking task represents the identification of the user's position. The user should be navigated to a specific location. The location can be declared as longitude and latitude attributes or as an address. The concrete navigation type can be a simple map, a compass or augmented reality-based 3D sign in the user's field of view. The interaction task is especially suitable to gain user-aware interactions. They are suitable to retrieve interactions to create checks e.g. retrieve a special interaction command to confirm the receiving of goods in the logistics domain. The environment identification task provides the declaration of the task in combination with an object identification e.g. for a package with a specific barcode some value-added-services have to be carried out. The information provision task allows the representation of information for the smart glasses user that is for instance necessary to execute complex tasks. The different kinds of information types could be an image, video, audio, checklist or FAQ. A picture or video-based task enables the recording of the environment e.g. pictures can be taken to collect defections in the process receipt of goods. As mentioned before the advanced communication task represents the communication possibilities of smart glasses for instance via e-mail, audio messages, video conferences or remote systems.

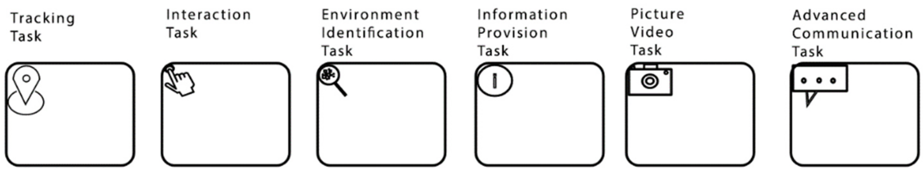


Figure 5. Activities of the BPMN Extension for Smart Glasses Applications (BPMN4SGA)

4.5 Implementation into a suitable Modelling Tool

To realize an application of the BPMN extension for smart glasses applications, we implemented the extension model into a modelling tool and extended the BPMN metamodel [24]. Business users should document smart glasses-based processes effortlessly without the need to install a new program. Therefore we favour web-based modelling tools over desktop-based modelling tools that need additional software modules e.g. language packets such as Java. To identify a suitable BPMN modelling tool that fulfills our requirements, we conducted a market analysis. Hence we do not use established desktop-based meta modelling platforms such as ADOxx [25] or MetaEdit+ [26]. Instead, we build upon BPMN-js a rendering toolkit for BPMN 2.0 that is based on JavaScript [24]. It allows a comfortable integration into web applications and is further extendable. The project is supported by the company Camunda and an active community. Further example projects simplify the customization and extension of the web-based modelling tool [27].

In figure 6 a screenshot from the adjusted modelling tool is presented. The left side shows a palette with the standard BPMN notation elements. The right side provides options to edit the element specific attributes. In the middle is the drawing surface. The domain-specific notation elements can be reached through a click on a standard BPMN element e.g. task or event. After that, a context menu opens and through a further click on the screw-wrench icon the BPMN element can be changed into a more specific form e.g. a start event into a voice start event. We represented BPMN4SGA elements in blue for precise identification. We do not include the concrete syntax in the palette (left side) to keep the clarity of the task/event notation elements and to retain the same user interface as for standard BPMN elements.

The icons are implemented into JavaScript code as SVG graphics. For a correct representation of the BPMN extension, the BPMN metamodel was enhanced with the domain-specific elements. This step ensured a correct transformation of the formal model into an XML document. Therefore, we edit the JSON sources that are representing the BPMN 2.0 metamodel. This is realized in bpmn.io through the bpmn-moddle library [28].

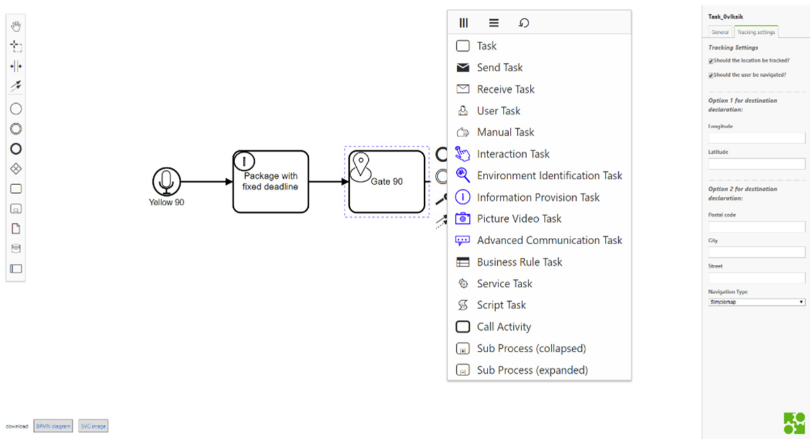


Figure 6. Demonstration of the BPMN4SGA in bpmn.io (screenshot)

5 Architectural Integration of the Model-driven Software Development Concept

MDSD aims to "automatically transforming formal models into executable code" [29]. This can be realized through the generation of code or interpretation. The generation of code is widely-used regarding structural parts, in contrast interpreters are used for behavior characteristics of a system [29]. Behavioral-oriented information system models are representing process and functional characteristics [30]. Because of, BPMN4SGA models focus on the representation of behavioral-oriented information systems, i.e. smart glasses applications, we favour the interpretation approach. Further reasons are the two differences between generators and interpreters that makes interpreters more suitable for our research project. The first reason is the time of analysis. The analysis of generators occurs during build time, whereas interpreter can parse models at runtime. Therefore, changes in BPMN4SGA models have a direct effect on smart glasses applications through direct interpretation. Second, the mode of execution differs. Interpreters execute directly different kind of code based on the read model. In contrast, generators link code fragments that are compiled in a next step [29]. Because the solution tries to contribute non-technical experts with a tool to readjust smart glasses applications directly (cf. section 2), we decided to implement an interpreter that binds code during runtime and does not need a further compilation step.

The architecture of our system is in accordance with Metzger et al. [11] and is shown in figure 7. The domain expert can model smart glasses-based processes through the web-modelling tool that includes the BPMN4SGA notation elements and attributes. The generated XML file can be uploaded through a web interface that saves metadata e.g. process id, process name and creation date from the process in the database. The web and REST interfaces for the platform are implemented with the PHP framework Symfony [31]. Further, an authentication system was implemented with Symfony that enables user-specific storing and client-side loading of BPMN4SGA models. Additionally, an Android PacKage (APK) was programmed and installed on the smart glasses. We implemented the generation system for the Vuzix M100. Thereby, the installed mobile app represents the generation system. The app includes the interpreter, which consists of an individual programmed parser that is implemented into the application. In doing so, the parser is written in Java and it is adjusted to our needs to read BPMN4SGA XML files. Based on the interpretations specific Android activities are called. We decided to download the XML files completely and execute the parsing steps locally on the smart glasses device to enable offline functionality and to minimize the number of requests. During the process execution, the generation system loads activity specific content through REST request e.g. pictures. Also, the system sends data to the server e.g. messages or log files and has further connections to additional systems e.g. to handle speech recognition.

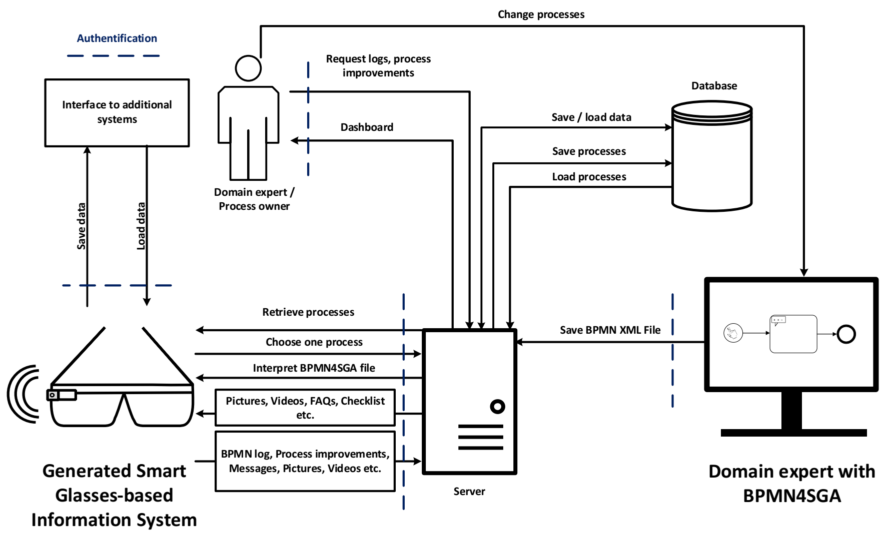


Figure 7. Architecture to generate Smart Glasses-based Information Systems with BPMN4SGA in accordance with Metzger et al. [11]

In close relation to figure 7 shows figure 8 the concrete MDSD approach: Thereby, a modelled process based on BPMN4SGA (1) will be interpreted by a parser (2) and calls specific views and methods. The triggered functions will be reflected in the user's field of view (3). The smart glasses use case "receipt of goods" is a common use case in the logistics domain [9] and is used to demonstrate the solution. The logistician starts the process with a specific speech command that can be specified in further attributes. Next, the worker has to scan the qrcode of the product to identify it. After that, he should give feedback regarding the condition of the goods. For instance, if the product is damaged, he creates a picture and does a damage classification. Finally, after the completion of the process tasks, the data is accessible and new processes can be triggered or selected by the smart glasses user [32].

The presented approach ensures high flexible information systems that can be changed by business experts without programming skills due to a direct transformation of the BPMN4SGA model into the desired information system. Thereby smart glasses-based information systems can in particular facilitate service processes [4].

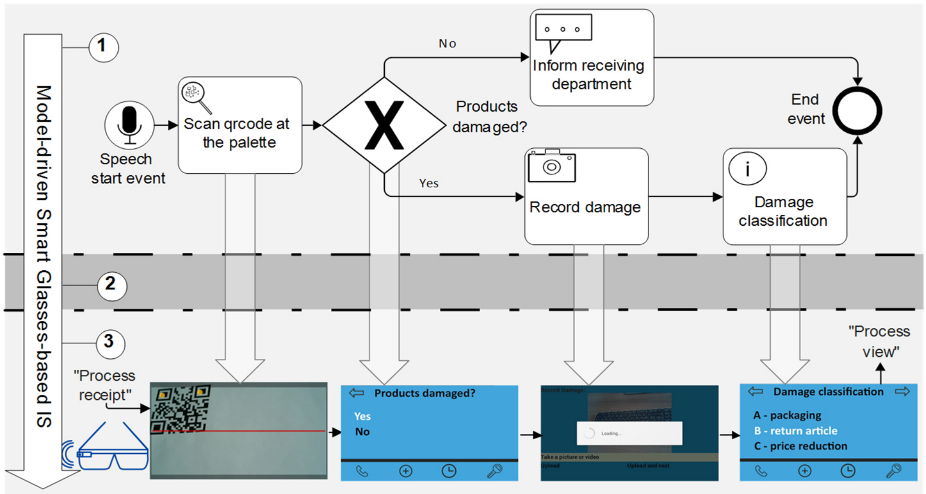


Figure 8. Demonstrating Smart Glasses-based Information System based on BPMN4SGA [32]

6 Discussion and Conclusion

Discussion. During the research project, some points for discussion arose. Additionally, we noticed limitations that are mentioned here. We are aware that the development and design of the BPMN4SGA are mostly based on the insights from the research project and our understanding of smart glasses applications. While we included requirements based on additional smart glasses literature, an overall requirement elicitation was not fully covered. In our point of view, the developed BPMN extension offers the possibilities to discuss with external experts and evaluate the implemented notation elements and attributes. Additional adjustments can be implemented in a further step. Moreover, we do not claim a full applicability of the developed generation system for every smart glasses-based use case – to cover most of them will be a further challenge in the future. Furthermore, the developed BPMN4SGA is mainly based on smart glasses functionalities and it is questionable if a broader software-based focus has to be included. For instance, the communication system was yet not implemented due to the complexity and further relationships with additional interfaces. We developed a closed ecosystem that is mainly focused on visualization purposes and does not reflect business boundaries e.g. to enterprise resource planning systems. So far the representation of businesses interfaces is underrepresented. Additionally, an evaluation with practice experts and field test can deliver valuable insights and are the next research steps.

Conclusion. This research project presents BPMN4SGA, the BPMN extension for smart glasses applications. The paper describes the technical development and implementation of the abstract and concrete syntax, and the implementation steps into a web-based modelling tool. The modelling tool is added into an architecture to realize smart glasses applications based on formal models through a MDSD approach. The implemented artifact enables the direct transformation of BPMN4SGA models into smart glasses applications. As technical DSML are still underrepresented in the range of various DSML, the research project contributes to the development and usage of these. Further, the research addresses a relevant topic to enhance the adoption of new technologies in businesses and delivers a solution to implement business specific smart glasses applications. The implemented solution enables non-technical domain experts to create and modify smart glasses applications for business use cases without the need to write one single code line.

7 Acknowledgements

This research is partly funded by the German Federal Ministry of Education and Research (BMBF) in the scope of the research project Glasshouse (ID: 02K14A090) and the Graduiertenkolleg va-eva: Vertrauen und Akzeptanz in erweiterten und virtuellen Arbeitswelten at the University of Osnabrueck. We would like to thank them for their support.

References

1. Klinker, K., Berkemeier, L., Zobel, B., Wüller, H., Huck-, V., Wiesche, M., Remmers, H., Thomas, O., Krcmar, H.: Structure for innovations : A use case taxonomy for smart glasses in service processes. In: Multikonferenz Wirtschaftsinformatik 2018, pp. 1599–1610. Lüneburg (2018)
2. Reif, R., Günthner, W.A., Schwerdtfeger, B., Klinker, G.: Pick-by-Vision comes on Age : Evaluation of an Augmented Reality supported Picking System in a real Storage Environment. In: Proccedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa, pp. 23–31. Pretoria (2009)
3. Streibl, M., Brandl, P.: Einsatz eines Live Video Remote Systems in der Industrie. In: Wyers, B. and Dittmar, A. (eds.) Mensch und Computer 2016, pp. 1–6. Gesellschaft für Informatik e.V., Aachen (2016)

4. Niemöller, C., Metzger, D., Thomas, O.: Design and Evaluation of a Smart-Glasses-based Service Support System. In: 13th International Conference on Wirtschaftsinformatik (WI), pp. 106–120. St. Gallen (2017)
5. Hobert, S., Schumann, M.: Enabling the Adoption of Wearable Computers in Enterprises – Results of Analyzing Influencing Factors and Challenges in the Industrial Sector. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 4276–4285. (2017)
6. Uchiyama, T., Tanuma, K., Fukuda, Y., Nakanishi, M.: GUI Design Solution for a Monocular, See-through Head-Mounted Display Based on Users' Eye Movement Characteristics. In: Shumaker, R. (ed.) Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments. VAMR 2013, LNCS, vol. 8021, pp. 231–240. Springer, Berlin Heidelberg (2013)
7. Clark, T., Barn, B.S.: Domain Engineering for Software Tools. In: Reinhartz-Berger, I., Sturm, A., Clark, T., Cohen, S., and Bettin, J. (eds.) Domain Engineering: Product Lines, Languages, and Conceptual Models, pp. 187–209. Springer, Berlin Heidelberg (2013)
8. Trompeter, J., Pietrek, G., Beltran, J.C.F., Holzer, B., Kamann, T., Kloss, M., Mork, S.A., Niehues, B., Thoms, K.: Modellgetriebene Softwareentwicklung - MDA und MDSD in der Praxis. entwickler.press, Frankfurt am Main (2007)
9. Niemöller, C., Zobel, B., Berkemeier, L., Metzger, D., Werning, S., Adelmeyer, T., Ickerott, I., Thomas, O.: Sind Smart Glasses die Zukunft der Digitalisierung von Arbeitsprozessen? Explorative Fallstudien zukünftiger Einsatzszenarien in der Logistik. In: Leimeister, J.M. and Brenner, W. (eds.) 13. International Conference on Wirtschaftsinformatik (WI), pp. 410–424. St.Gallen (2017)
10. Vogel, J., Zobel, B., Jannaber, S., Thomas, O.: BPMN4SGA: A BPMN Extension for Smart Glasses Applications to enable Process Visualisations. In: Czarnecki, C., Brockmann, C., Sultanow, E., Koschmider, A., and Selzer, A. (eds.) Workshops der Informatik 2018 - Architekturen, Prozesse, Sicherheit und Nachhaltigkeit, pp. 259–273. Köllen Druck+Verlag GmbH, Bonn (2018)
11. Metzger, D., Niemöller, C., Berkemeier, L., Brenning, L., Thomas, O.: Vom Techniker zum Modellierer – Konzeption und Entwicklung eines Smart Glasses Systems zur Laufzeitmodellierung von Dienstleistungsprozessen. In: Smart Service Engineering, pp. 193–213. Springer Fachmedien Wiesbaden, Wiesbaden (2017)
12. Fellmann, M., Metzger, D., Jannaber, S., Zarvic, N., Thomas, O.: Process Modeling Recommender Systems. Bus. Inf. Syst. Eng. 60, 21–38 (2018)
13. Petersen, N., Pagani, A., Stricker, D.: Real-time modeling and tracking manual workflows from first-person vision. In: 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 117–124. IEEE (2013).
14. Oesterreich, T.D., Teuteberg, F.: Why one big picture is worth a thousand numbers: measuring intangible benefits of investments in augmented reality based assistive technology using utility effect chains and system dynamics. Inf. Syst. E-bus. Manag. 16, 407–441 (2018)
15. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 24, 45–77 (2008)
16. Braun, R., Schlieter, H.: Requirements-based development of BPMN extensions: The case of clinical pathways. In: 2014 IEEE 1st International Workshop on the Interrelations between Requirements Engineering and Business Process Management (REBPM), pp. 39–44. IEEE, Karlskrona, Sweden (2014)

17. Stroppi, L.J.R., Chiotti, O., Villarreal, P.D.: Extending BPMN 2.0: Method and Tool Support. In: Dijkman, R., Hofstetter, J., and Koehler, J. (eds.) Business Process Model and Notation. BPMN 2011, LNBIP, vol. 95, pp. 59–73. Springer, Berlin Heidelberg (2011)
18. Object Management Group (OMG): Business Process Model and Notation (BPMN), http://www.omg.org/spec/BPMN/2.0.2/PDF/ (Accessed: 20.12.2018).
19. Yousfi, A., Bauer, C., Saidi, R., Dey, A.K.: UBPMN: A BPMN extension for modeling ubiquitous business processes. Inf. Softw. Technol. 74, 55–68 (2016)
20. Braun, R., Schlieter, H., Burwitz, M., Esswein, W.: BPMN4CP: Design and Implementation of a BPMN Extension for Clinical Pathways. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 9–16. IEEE (2014)
21. Frank, U.: Domain-Specific Modeling Languages: Requirements Analysis and Design Guidelines. In: Reinhartz-Berger, I., Sturm, A., Clark, T., Cohen, S., and Bettin, J. (eds.) Domain Engineering: Product Lines, Languages, and Conceptual Models, pp. 133–157. Springer, Berlin Heidelberg (2013)
22. NounProject, https://thenounproject.com/ (Accessed: 20.12.2018).
23. Vuzix M300 Technical Specifications, https://www.vuzix.com/products/m300-smart-glasses (Accessed: 20.12.2018).
24. Bpmn.io: BPMN 2.0 rendering toolkit and web modeler., https://bpmn.io/toolkit/bpmn-js/ (Accessed: 20.12.2018)
25. Fill, H., Karagiannis, D.: On the Conceptualisation of Modelling Methods Using the ADOxx Meta Modelling Platform. Enterp. Model. Inf. Syst. Archit. 8, 4–25 (2013)
26. Kelly, S., Lyytinen, K., Rossi, M., Tolvanen, J.P.: MetaEdit+ at the Age of 20. In: Bubenko, J., Krogstie, J., Pastor, O., Pernici, B., Rolland, C., and Sølvberg, A. (eds.) Seminal Contributions to Information Systems Engineering: 25 Years of CAiSE, pp. 131–137. Springer, Berlin Heidelberg (2013)
27. BPMN-JS-Examples, https://github.com/bpmn-io/bpmn-js-examples (Accessed: 20.12.2018)
28. BPMN-MODDLE, https://github.com/bpmn-io/bpmn-moddle (Accessed: 20.12.2018)
29. Stahl, T., Völter, M.: Model-Driven Software Development. John Wiley & Sons, Southern Gate, Chichester (2006)
30. Schütte, R.: Grundsätze ordnungsmäßiger Referenzmodellierung. Gabler Verlag (1998).
31. Symfony The Book, https://symfony.com/pdf/Symfony_book_2.8.pdf (Accessed: 20.12.2018)
32. Vogel, J., Jannaber, S., Zobel, B., Thomas, O.: Design and Development of a Process Modelling Environment for Business Process Utilization within Smart Glasses. In: Fellmann, M. and Sandkuhl, K. (eds.) 9th International Workshop on Enterprise Modeling and Information Systems Architectures (EMISA), pp. 85–89. Rostock (2018)

Using Blockchain in Peer-to-Peer Carsharing to Build Trust in the Sharing Economy

Paul Bossauer[1], Thomas Neifer[1], Christina Pakusch[1], and Paul Staskiewicz[1]

[1] Bonn-Rhein-Sieg University of Applied Sciences, Department of Management Sciences, Sankt Augustin, Germany
{paul.bossauer,thomas.neifer,christina.pakusch}@h-brs.de,
paul.staskiewicz@smail.wis.h-brs.de

Abstract. Trust is the lubricant of the sharing economy, especially in peer-to-peer carsharing where you leave a valuable good to a stranger in the hope of getting it back unscathed. Central mechanisms for handling this information gap nowadays are ratings and reviews of other users. The rising of connected car technology opens new possibilities to increase trust by collecting and providing e.g. driving behavior data. At the same time, this means an intrusion into the privacy of the user. Therefore, in this work we explore technological approaches that allow building trust without violating the privacy of individuals. We evaluate to what extent blockchain technology and smart contracts are suitable technologies to meet these challenges by setting up a prototype implementation of a blockchain-based carsharing approach. In this context, we present our research approach and evaluate the prototype in terms of trust and privacy.

Keywords: Blockchain, Carsharing, Trust, Peer-to-Peer, Privacy

1 Introduction

The sharing economy market has grown strongly in recent years [1, 2]. Peer-to-peer platforms (short: P2P-platforms) such as AirBnB, UBER, Drivy or BlaBlaCar enable private individuals to make a good or a service (e.g. a room, a trip or a car) available to someone else for a certain period of time or to use this good for a fee. Looking at P2P-carsharing, there is – in addition to the fee – a difference between the sharing of a car e.g. within the family or with a complete stranger. Therefore, trust plays a central role here [3–6]. Trust is also sometimes referred to as the *currency* or *lubricant* of the sharing economy, as it is an efficient way of reducing search and informational effort in social exchanges [7].

The lending of cars is a particularly emotional topic for many people. Traditional sharing concepts are often based on reputation systems where users evaluate each other. The rating and review processes aim at establishing trust towards unknown persons. The evaluation in such reputation systems is usually characterized by subjective ratings and reviews in respect of the experiences made during the rental process. The traditional reputation system in P2P-carsharing often reaches its limits, so that tenants often have a lack of confidence in a neutral damage event regulation [8]. Due to these challenges,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

P2P-carsharing is still a niche market in the everyday mobility. In this paper we want to examine, how blockchain can help to reduce these barriers and build trust by creating a blockchain-based P2P-carsharing prototype and evaluating it in existing living lab environments.

2 Related Work

A relatively new development in carsharing is the growth of P2P-carsharing platforms such as Turo, Snappcar, Getaway or Drivy, which allow users to rent or lend cars to each other. In contrast to the local context of traditional carsharing, the peers usually do not know each other personally. This gives rise to a similar challenge as in the case of e-commerce and online auctions to build trust among strangers via the Internet [9]. Over the last decades, reputation systems have been adopted as a prevailed mechanism, where users leave comments and rate others. This source of information was accomplished by information provided by oneself like name, images, and personal preferences. Both types of information serve as surrogates to build trust in an online world [10].

Nowadays, computational gathered information using connected car technology could be used as an additional source for building trust, e.g. monitoring the driving behavior and calculate trust scores based on these data. This, however, raises new questions with regard how these sources of information are valued by the users to gain knowledge about the other in order to increase the willingness to offer her or his good or service. In addition, it also raises new privacy issues as users must balance the costs to disclose the privacy opposite to the benefits of a better reputation.

There is no consistent definition of trust in the literature, but it is generally understood as a multidimensional, socio-psychological construct [3, 7]. Hawlitschek et al. [3] see trust as the expectation and obligation that an exchange will take place in the future. In this paper we follow the definition of Huurne et al. [7]:

"[Trust is] the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party"

This type of trust is particularly important in potentially risky and insecure situations where parties are interdependent [11]. Such situations are typical for the sharing economy since Internet-based mediation removes the usual mechanisms for developing social and economic bonds that promote the emergence of trust. In addition to the trust in the peer, trust in the platform and the product made available must also be built [3].

To solve this trust problem, an often-discussed technology is the blockchain [12–18]. Miraz describes Blockchain as the "Trust Machine", as it preserves a permanent record of all transactions while making sure that any identity-related information of users can be kept incognito [17]. Weber et al. [18] used the Blockchain to monitor and execute untrusted business processes with smart contracts. They explored the fundamental problem of trust in collaborative processes by using smart contracts in three use cases: supply chain choreography, incident management choreography, and insurance claim handling. There are various studies exploring the impact of blockchain on trust, but none in the context of P2P-carsharing. Sarantini et al. give an insight how blockchain-based mobility concepts may look like in the future, but without discussing the trust issue in the sharing economy [19]. Therefore, we want to address this gap and take a closer look at how capable Blockchain is in solving trust issues in P2P-carsharing.

3 Research approach

In the first step, we carried out three focus group interviews and eight problem-centered individual interviews *to understand how car-lenders and car-tenants deal with connected car data and derive implications for blockchain-based carsharing*. We explored how these new possibilities of computational trust mechanisms are perceived by the users and discussed, what sensors might be relevant and what information they would be personally interested if they were a lender as well as where they see privacy concerns from the perspective of a tenant. In the focus group interviews the participants evaluated the data and information discussed from the perspectives of lenders and tenants, whereas in the individual interviews only one perspective was asked. The decision to analyze only one perspective in the individual interviews was made deliberately to prevent distortion of the results due to a perspective transfer. The focus groups were divided into mobility professionals (3 women, 3 men, 20-48 years), those interested in innovative sharing concepts (1 woman, 5 men, 20-32 years) and those highly interested in private car sharing (3 men, 28-47 years). The problem-centred individual interviews were divided between potential tenants and lenders (1 woman and 3 men per group, 20-65 years).

In the second step, we are programming a prototype in the sense of a minimum viable product (MVP) [20] for a blockchain-based carsharing platform. By iteratively testing, evaluating and improving the prototype in close cooperation with the users, we consider the Action Design Research principles [21]. The goal of the first prototype is *to achieve that users can execute the basic functions to get a fast feedback*. These include the initial rent, the return, and the payment. Each car creates a deployed contract on the Ethereum blockchain. Regarding to the MVP principle, we decided to use the Ethereum Blockchain, due to the high distribution as well as many existing projects and tutorials. When the prototype reaches an advanced stage of development, we must evaluate existing technologies in detail and decide whether a purpose-built or derived blockchain is more advantageous. Once the basic functionalities have been integrated, we want to add new functionalities to the smart contracts to simulate the scenario of the allocation of data to specific events.

The third step of our research will be the evaluation of the prototype in a living lab. In this step, *the hypothesis shall be tested whether the blockchain technology and the associated transparency of data access can increase the trust on the part of lenders and tenants*. The carsharing process will be carried out and validated via interviews and app feedback in a real environment with test households. Together with the participants, the prototype is to be further developed and evaluated according to their requirements.

4 Discussion and Conclusion

The current state of our research shows that many people do not use P2P-carsharing, because they are not ready to share their car with strangers due to a lack of trust. The lack of trust is particularly evident in the interviews with potential car lenders. From the tenant's perspective, privacy is often the reason for not disclosing their data. However, our findings show that the privacy concerns depend on the appropriateness and proportionality of the purposes. The willingness to respect the principle of reciprocity increases user acceptance and confidence that the data will not be misused. In this regard, respondents frequently argued that they accept the disclosure of data under certain conditions, such as e.g. accidents or the crossing of a spatial boundary. With smart contracts, such conditions can be technically implemented and thus allow a higher security for car lenders in cases of emergency, as well as the protection of privacy for car tenants if the emergency does not occur.

Currently, we are analyzing the findings from the focus groups and the problem-centered individual interviews more deeply to derive requirements for the prototype of the blockchain-based carsharing platform. Especially the data disclosure in certain situations or after certain events will be a challenge in the programming of smart contracts. Since these requirements differ from user to user, there must be a possibility that car lenders and car tenants can negotiate about the disclosure of certain types of data. After the requirements have been implemented in the prototype, it is to be researched and further developed in a real environment with a vehicle. In the literature, blockchain is often described as a substitute for intermediation [22]. An essential role of intermediaries in the internet economy is the regulation of supply and demand and in the sharing economy especially the building of trust [23]. Only by evaluating our prototype in a real environment will it become clear, whether blockchain technology plays a key role in building trust in the sharing economy or not.

References

1. PricewaterhouseCoopers: Share Economy in Deutschland wächst weiter, https://www.pwc.de/de/pressemitteilungen/2018/share-economy-in-deutschland-waechst-weiter.html.
2. Zervas, G., Proserpio, D., Byers, J.W.: The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. J. Mark. Res. 54, 687–705 (2017).
3. Hawlitschek, F., Teubner, T., Weinhardt, C.: Trust in the sharing economy. Unternehm. 70, 26–44 (2016).
4. Belk, R.: Sharing. J. Consum. Res. 36, 715–734 (2009).
5. Botsman, R., Rogers, R.: What's mine is yours: how collaborative consumption is changing the way we live. (2011).
6. Botsman, R.: Who Can You Trust?: How Technology Brought Us Together–and Why It Could Drive Us Apart. Penguin UK (2017).
7. Huurne, M., Ronteltap, A., Corten, R., Buskens, V.: Antecedents of trust in the sharing economy: A systematic review. J. Consum. Behav. 16, 485–498 (2017).

8. Stevens, G., Bossauer, P., Jakobi, T., Pakusch, C.: Mehrseitiges Vertrauen bei IoT-basierten Reputationssystemen. Mensch Comput. 2018-Work. (2018).
9. Resnick, P., Zeckhauser, R.: Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. In: The Economics of the Internet and E-commerce. pp. 127–157. Emerald Group Publishing Limited (2002).
10. Wang, Y.D., Emurian, H.H.: An overview of online trust: Concepts, elements, and implications. Comput. Hum. Behav. 21, 105–125 (2005).
11. McKnight, D.H., Chervany, N.L.: What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. Int. J. Electron. Commer. 6, 35–59 (2001).
12. Dorri, A., Kanhere, S.S., Jurdak, R.: Towards an optimized blockchain for IoT. In: Proceedings of the Second International Conference on Internet-of-Things Design and Implementation. pp. 173–178. ACM (2017).
13. Dorri, A., Steger, M., Kanhere, S.S., Jurdak, R.: Blockchain: A distributed solution to automotive security and privacy. IEEE Commun. Mag. 55, 119–125 (2017).
14. Dorri, A., Kanhere, S.S., Jurdak, R., Gauravaram, P.: Blockchain for IoT security and privacy: The case study of a smart home. In: Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on. pp. 618–623. IEEE (2017).
15. Kosba, A., Miller, A., Shi, E., Wen, Z., Papamanthou, C.: Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In: 2016 IEEE symposium on security and privacy (SP). pp. 839–858. IEEE (2016).
16. Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., Njilla, L.: Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In: Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. pp. 468–477. IEEE Press (2017).
17. Miraz, M.H.: Blockchain: Technology Fundamentals of the Trust Machine. Mach. Lawyering Chin. Univ. Hong Kong 23rd Dec. (2017).
18. Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., Mendling, J.: Untrusted business process monitoring and execution using blockchain. In: International Conference on Business Process Management. pp. 329–347. Springer (2016).
19. Saranti, P.G., Chondrogianni, D., Karatzas, S.: Autonomous Vehicles and Blockchain technology are shaping the future of Transportation.
20. Ries, E.: The lean startup: how constant innovation creates radically successful businesses. Portfolio Penguin, London (2011).
21. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS Q. 37–56 (2011).
22. Tapscott, D., Tapscott, A.: How blockchain will change organizations. MIT Sloan Manag. Rev. 58, 10 (2017).
23. Clement, R., Schreiber, D.: Internet-Ökonomie. Springer (2016).

Testing in Big Data: An Architecture Pattern for a Development Environment for Innovative, Integrated and Robust Applications

Daniel Staegemann[1], Johannes Hintsch[1] and Klaus Turowski[1]

[1] Otto-von-Guericke University, Faculty of Computer Science, Magdeburg, Germany
{daniel.staegemann,johannes.hintsch,klaus.turowski}@ovgu.de

Abstract. Big Data is a crucial pillar for many of today's newly emerging business models. Areas of application range from consumer analysis over medicine to fraud detection. All of those domains require reliable software. Even though imperfect results are accepted in Big Data software, bugs and other defects can have drastic consequences. Therefore, in this paper, the software engineering sub discipline of testing is addressed. Big Data exhibits characteristics which differentiate its processing software from those that process traditional workloads. Consequently, an architecture pattern for testing that can be integrated into development environments for Big Data software is proposed. The paper features a detailed description of the artifact as well as a preliminary plan for evaluation.

Keywords: Big Data, Testing, Design Science, Software Engineering.

1 Introduction

Big data ranked the top-most important area of IT investments throughout the past five consecutive years [1]. Firms use data to get new insights (e.g., about customers' purchasing preferences) or to make decisions (e.g., in credit card fraud management). Even though the potential is high [2], companies are struggling to cope with the implicated challenges [3–5]. As an important part of the software development process "Software testing is a process, or a series of processes, designed to make sure computer code does what it was designed to do and, conversely, that it does not do anything unintended" [6]. Therefore, all activities that are supposed to determine the congruence of a program and its pre-defined requirements can be deemed software testing. The necessity to rigorously test software stems from the potential harm, that even seemingly little mistakes in the software can cause [7]. Architectures are the "fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution" [8]. Big Data, as a new paradigm, challenges the architecture of traditional software engineering environments, particularly in testing [9]. This is due to the properties of Big Data, often characterized as the four "V"s. Those are volume (amount of data), variety (different sources of data), velocity (rate of the dataflow) and variability

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

(change of data characteristics). These characteristics overstrain traditional data architectures and require new techniques, like the usage of horizontal scaling, to efficiently handle the respective datasets. Those challenges are also reflected in the related testing necessities [10].

To accommodate those necessities we follow the Design Science Research [11] paradigm to outline a testing architecture to support the development of big data applications. The focus is on domain specific applications that facilitate investigating the meaning of data and the relationships between different data. So far, social media analysis is deemed a promising domain for investigation.

2 Related Work

The diversity of different preliminary works in the existing literature reflects the complexity as well as the relevancy of the topic. It ranges from general descriptions of problem areas that also require testing [12], concepts on how to benchmark or test in the area of Big Data [13–16] and on the challenges of quality assurance [17] to more concrete approaches like an implementation for dataless testing [18]. There is however not a universally optimal solution yet, resulting in a need for further research.

3 Artifact

As mentioned beforehand, there are significant differences between software solutions for traditional data and those for Big Data. This results in additional challenges that need to be considered in the testing process as well as in the corresponding architecture of a software engineering environment. The three most notable challenges for testing in Big Data are the following. These were derived from literature [19–22] and from discussions with two experts.

- Difference 1: In contrast to traditional software in Big Data applications non-functional properties (like the ability to handle high volume and velocity) have a higher importance [10].
- Challenge 1: Huge amounts of varying data are required to test non-functional properties.
- Difference 2: Data are often heterogeneous (variety, variability) and the data quality is often poor [16].
- Challenge 2: Necessity to test the clearing and converting of source data.
- Difference 3: Due to the use-cases there is often a higher difficulty to determine if the tested system delivers optimal results [17].
- Challenge 3: The system is drafted to tackle situations that are complex in terms of data and could therefore not be handled with traditional technology, for this reason there is often no known set of inputs and matching outputs. (oracle problem) [23].

To provide maximum value, an architecture for testing Big Data systems should offer solutions for all of the mentioned challenges. Since the reviewed existing approaches were considered not sufficient in light of those challenges, the proposed one was created from scratch.

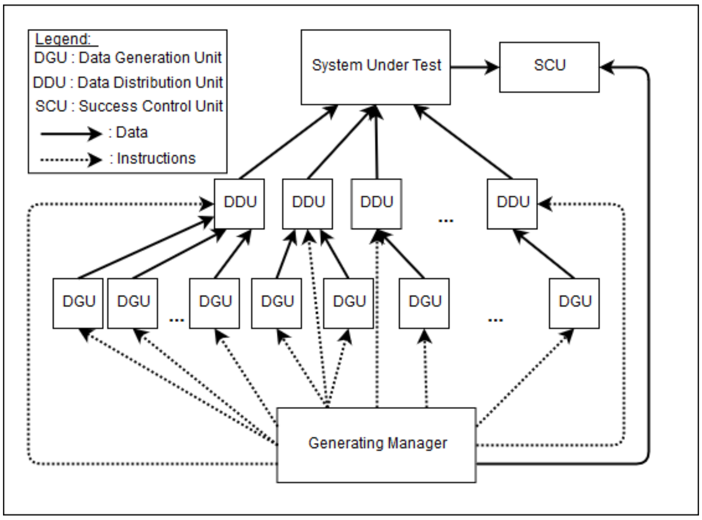


Figure 1. Architecture pattern for testing in Big Data development environments

The proposed architecture pattern, shown in Figure 1, consists of several elements, that aim to fulfill the identified requirements, when combined, to extensively stress the System under Test (SUT). A Generating Manager (GM) controls the whole process and steers the Data Generating Units (DGU) as well as the Data Distribution Units (DDU). If needed the GM can also create and terminate DGUs and DDUs. It is currently investigated, if an algorithm based on MapReduce [24] might be suitable for organizing the test procedure. The Success Control Unit (SCU) monitors the test, comparing the information sent by the GM with the results of the SUT. This allows for a real time monitoring of the performance of the SUT, regarding functional as well as non-functional aspects. The DGUs are each specialized on outputting one type of data (e.g. Twitter posts, reviews) and if needed specific characteristics (e.g. incompleteness, conflicting statements). This allows to choose the best possible solution for each creation sub-task instead of being bound to a solution that delivers acceptable but possibly suboptimal test data for all cases. This approach aims at testing the SUT's clearing, converting and processing of source data by feeding it data of varying type and quality, therefore tackling challenge 2.

Each DGU can generate data from scratch, by recreating existing data patterns, or outputs data that are provided by existing databases or data scientists. For this purpose it is given instructions by the GM. It is possible to have several DGUs with the same characteristics to achieve a higher rate of data generation. It is also feasible to create DGUs that are only providing data corresponding to the pattern the SUT is supposed to detect, while other DGUs are creating "decoy data" that does not comply. The chosen approach targets an easy assessment regarding the detection rate of the SUT. This is because the GM knows which DGU's data are supposed to be detected by the SUT, therefore enabling the SCU as a test-oracle, addressing challenge 3. The DDUs are each devoted to one type of data, therefore utilizing the possible benefits of specialization. They are forwarded the data directly by the DGUs assigned to them by the GM. In the DDUs a buffer of data can be created for further use. When ordered by the GM, the DDUs send their data to the SUT, using the requested pattern, volume and velocity, utilizing the buffered data if needed, taking on challenge 1.

4 Evaluation

The evaluation follows the pattern proposed by Sonnenberg and vom Brocke [25]. EVAL 1 explores if the research and the accompanying creation of an artifact are justified or unnecessary. This step is included in the publication at hand. The general need for research in the outlined topic was illustrated, experts and relevant literature were included in the derivation of significant challenges and those were subsequently the foundation of the taken design decisions. This results in the hypothesis that the proposed architecture constitutes an improvement compared to existing approaches. EVAL 1 can therefore be deemed as completed. EVAL 2 focuses the feasibility and practicability of the suggested approach. It will use logical reasoning, comparing the challenges and the solutions, provided by the artifact, as well as an analysis to verify if the chosen test organization algorithm terminates and expert interviews, e.g. concerning expectable performance, to judge the feasibility of the developed architecture and to remedy possible flaws in the architecture or the algorithm. The prototype of the artifact itself and its testing constitute EVAL 3. Once the concept is implemented in real-life scenarios, a case study and further expert interviews are planned (EVAL 4). An overview of these described steps is depicted in Table 1.

Table 1. Evaluation Plan

Evaluation Steps	*Description*	*Status*
EVAL 1	This publication	Completed
EVAL 2	Logical reasoning and expert interview	Planned
EVAL 3	SAP HANA and OpenStack based prototyping	Planned
EVAL 4	Case study and expert interview	Planned

5 Conclusion

Big Data poses new challenges compared to traditional software engineering. The same applies to the corresponding testing. As a consequence there is currently no universally applied approach for testing Big Data systems. Using the modular artifact introduced in this publication provides possible solutions for those challenges of testing Big Data applications, while still respecting the potential uniqueness of individual projects and the belonging test scenarios.

References

1. Kappelman, L., Johnson, V., Maurer, C., McLean, E., Torres, R., David, A., Nguyen, Q.: The 2017 SIM IT Issues and Trends Study. MIS Quarterly Executive 17, 53–88 (2018)
2. McAfee, A., Brynjolfsson, E.: Big Data: The Management Revolution: Exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision-making culture. Harvard Business Review 91, 1–9 (2012)
3. Lee, I.: Big data: Dimensions, evolution, impacts, and challenges. Business Horizons 60, 293–303 (2017)
4. Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F.: Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth 10, 13–53 (2017)
5. Tiwari, S., Wee, H.M., Daryanto, Y.: Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering 115, 319–330 (2018)
6. Myers, G.J., Badgett, T., Sandler, C.: The art of software testing. J. Wiley & Sons, Hoboken, N.J (2011)
7. Patton, R.: Software testing. SAMS, Indianapolis (2001)
8. ISO/IEC/IEEE Systems and software engineering -- Architecture description. IEEE, Piscataway, NJ, USA 42010:2011
9. Otero, C.E., Peter, A.: Research Directions for Engineering Big Data Analytics Software. IEEE Intell. Syst. 30, 13–19 (2015)
10. NIST: NIST Big Data Interoperability Framework: Volume 1, Definitions. National Institute of Standards and Technology (2015)
11. Hevner, A., R, A., March, S., T, S., Park, Park, J., Ram, Sudha: Design Science in Information Systems Research. Management Information Systems Quarterly 28, 75-105 (2004)
12. Jin, X., Wah, B.W., Cheng, X., Wang, Y.: Significance and Challenges of Big Data Research. Big Data Research 2, 59–64 (2015)
13. Mahesh Gudipati, S. Rao, Naju D. Mohan, Naveen Kumar Gajja: Big Data : Testing Approach to Overcome Quality Challenges. Infosys Labs Briefings 11, 65–73 (2013)
14. Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S., Buyya, R.: Big Data computing and clouds: Trends and future directions. Journal of Parallel and Distributed Computing 79-80, 3–15 (2015)
15. Han, R., Lu, X., Xu, J.: On Big Data Benchmarking. In: Zhan, J., Han, R., Weng, C. (eds.) Big Data Benchmarks, Performance Optimization, and Emerging Hardware, vol. 8807, pp. 3–18. Springer International Publishing, Cham (2014)
16. Alexandrov, A., Brücke, C., Markl, V.: Issues in big data testing and benchmarking. In: Sixth International Workshop on Testing Database Systems - DBTest '13, pp. 1–5. ACM Press, New York (2013)
17. Tao, C., Gao, J.: Quality Assurance for Big Data Applications– Issues, Challenges, and Needs. In: The Twenty-Eighth International Conference on Software Engineering and Knowledge Engineering, pp. 375–381. KSI Research Inc. and Knowledge Systems Institute Graduate School, Pittsburgh (2016)
18. Ashoke, S., Haritsa, J.R.: CODD: A dataless approach to big data testing. Proceedings of the VLDB Endowment 8, 2008–2011 (2015)
19. Fan, J., Han, F., Liu, H.: Challenges of Big Data Analysis. National science review 1, 293–314 (2014)

20. Yang, A., Troup, M., Ho, J.W.K.: Scalability and Validation of Big Data Bioinformatics Software. Computational and structural biotechnology journal 15, 379–386 (2017)
21. Alyass, A., Turcotte, M., Meyre, D.: From big data analysis to personalized medicine for all: challenges and opportunities. BMC medical genomics 8, 33 (2015)
22. Kambatla, K., Kollias, G., Kumar, V., Grama, A.: Trends in big data analytics. Journal of Parallel and Distributed Computing 74, 2561–2573 (2014)
23. Barr, E.T., Harman, M., McMinn, P., Shahbaz, M., Yoo, S.: The Oracle Problem in Software Testing: A Survey. IIEEE Trans. Software Eng. 41, 507–525 (2015)
24. Dean, J., Ghemawat, S.: MapReduce. Communications of the ACM 51, 107–113 (2008)
25. Sonnenberg, C., Vom Brocke, J.: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B. et al. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice, vol. 7286, pp. 381–397. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)

Track 4:

Lern- und Wissensmanagement

Track Chairs:

Prof. Dr. Matthias Schumann
University of Goettingen

Prof. Dr. Ronald Maier
University of Innsbruck

eGovernment Competences revisited – A Literature Review on necessary Competences in a Digitalized Public Sector

Bettina Distel[1], Nadine Ogonek[1] and Jörg Becker[1]

[1] University of Münster - ERCIS, Chair for Information Systems and Information Management, Münster, Germany
{bettina.distel,nadine.ogonek,joerg.becker}@ercis.uni-muenster.de

Abstract. With the growing proliferation of digital technologies, organizations at all levels are faced with a changing environment to which employees and leaders have to adapt. Digital competences can be considered a key factor for the successful implementation of digital technologies in organizations of all kinds. Public administrations are exemption from this trend and are the focus in this study as a special organizational form. Despite the importance of these competences, extensive research on this subject is yet missing. To better understand the shape of knowledge contributions made so far, a structured literature review is conducted to uncover the state of the art of research on these competences. Results of this study indicate that only very few scholars have so far researched public administration competences more closely. Besides functional competences, a focus on so-called soft skills and personality traits is unveiled. Based on the findings, directions for future research are derived.

Keywords: eGovernment, digitalization, competences, literature review

1 Introduction

"[P]olicy makers face a race between technology and education, and the winners will be those who encourage skill upgrading so that all can benefit from digital opportunities" [1]. This statement nowadays becomes a postulate amongst decision makers worldwide, given the growing realization that digitalization is not just a temporary phenomenon but rather a revolutionary game-changing intrusion: "The number of internet users has more than tripled in a decade—from 1 billion in 2005 to an estimated 3.2 billion at the end of 2015. This means that businesses, people, and governments are more connected than ever before" [1]. On the one side, this offers great potentials, but at the same time organizations' and peoples' competences are heavily challenged by digitalization, needing to constantly adapt to an ever changing environment [2]. This is equally true for all organizations that in large part have acknowledged the need for the right skills in a digital world, but when being asked if

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

their employees embody these competences “[…] 53% rather or strongly disagreed with the statement” [3]. The increased use of information systems has entered public administrations worldwide equally and governments also have started to recognize the potential benefits in terms of efficiency and effectiveness gains, if comprehensively leveraged. Yet, “[s]uccessful digital transformation comes not from implementing new technologies but from transforming an organisation to take advantage of the possibilities that new technologies provide” [4]. Thus, the transformation needs to be thoroughly prepared and implemented, since it greatly affects the way the work is done and organized, and requires organizational adjustments [5], [6]. [7] point to the fact that employees from all organizational layers need to refresh their digital skills as a prerequisite for digital transformation to be successful. [8] advocate for a better understanding of this changed environment in order to equip “[…] students with useful models, methods, and technical skills for customer-centric and service-centric information systems” [8].

The basic assumption underlying this study is that competences are a key resource for the successful implementation and use of digital technologies. If public administrations wish to exploit the full potential of digitalization, the development of competences to deal with a diverse set of technologies should therefore be of equal importance as the development of the technology itself. Yet, in spite of the topic’s importance, scientific contributions in this area, are only rarely to be found and do not seem to be among the domain’s research priorities [9], [10], [11]. There are few attempts that started the endeavor to structure the competence necessities for eGovernment education, e.g. [12], [13], [14]. In this context, it is especially [14], who are the only ones that offer a differentiated clustering of necessary competences. Their study, however, was already published in 2015. Therefore, we would like to find out, how this line of research has developed since then and put forward the research question: *Which competences do public administrations employees need from a research perspective?*

2 Research Background

Next to providing definitions of key concepts in order to establish a clear understanding, we also investigate existing research on competences. The term *competence* used throughout this study refers to the combination of an individual’s work-related knowledge, skills and abilities [15]. In accordance with prior research [16], we propose that competences are a key prerequisite for the skillful use of information systems. With the increased use of Information Systems (IS), the way public bodies act is and will keep changing considerably. In general, the organization’s technical infrastructure has the power to improve its performance significantly. "In particular, information systems are considered to be a major asset for leveraging organizational transformation owing to the disruptive nature of IT [Information Technology] innovations, the deep digitalization of business, and their cross-organization and systemic effects, notwithstanding the amounts of investments in enterprise systems." [17]

Even though there are huge efficiency and productivity gains ascribed to the use of IS in public administrations [18], employees have to be provided with digital competences to be able to work in the digital age [19]. The adequate preparation of public servants is therefore indispensable, because the success of implementing IT highly depends on the employees' skills and expertise [20], [21]. Despite its importance, as has been mentioned before, education in the area of eGovernment has not yet been a major topic of interest in research. However, there have been limited intents to identify the necessary competences. Early studies on required competences in the public sector mainly focused on IT competences, e.g. [2], [22]. It is rather later studies that acknowledge the need for more diversified skill sets [23], [24]. Yet, it seems that these competences are not very well covered. According to a report by the European Commission, almost one quarter (23%) of the total EU population has no digital competences at all. Although this ratio is better amongst the working population, where only 14% have no digital skills at all, 39% are considered to have an insufficient level of digital skills [25].

However, quite some practice frameworks have been set up to classify and cluster the necessary competences, notably in the context of the ongoing *digitalization*, which can be referred to as "[…] the manifold sociotechnical phenomena and processes of adopting and using these technologies in broader individual, organizational, and societal contexts." [8] The Skill Framework for the Information Age (SFIA), for instance, is a reference guide, describing 97 skills for employees in information systems-related roles of any type. It provides a reference model embracing two dimensions, namely skills and different levels of responsibility [26]. Another framework to cluster "e"-related competences is the European e-competence framework (e-CF) [27]. The framework, initially developed in 2005, is designed as a means for describing necessary IT professionals' skills and knowledge requirements. It is based on 40 predefined competences, split into five ICT (Information and Communication Technology) areas. It relates to the European Qualifications Framework with five different proficiency levels. Both frameworks offer guidance for practitioners working with IT, however they do not consider the peculiarities of the public sector, inhibiting their application to the domain of eGovernment.

Therefore, another study, developed by [14], seems to be more promising, since it offers a categorization of different competence categories and is especially designed for the public sector. It builds on the findings from earlier studies in the same domain. Consisting of five different competence categories, which are composed of technical, socio-technical, organizational, managerial and political-administrative competences, it offers a more differentiated view on this topic. Technical competences are IT-related skills like the fundamentals, strategy and design of Information Systems (IS). Socio-technical competences encompass all the skills that are located at the interface of technical systems and human beings and involve both of them. Examples for such competences are framework requirements on the impact of eGovernment/technology. Organizational competences refer to the organizational integration of IT/eGovernment, organizational structures, process management etc. The category of managerial competences deals with business-related and management skills in the context of IT/eGovernment, like project, change and financial management. Political-administrative competences consist of skills that deal with the environment that

IT/eGovernment is embedded in. Examples are legal conditions and policies. These competences seem to better grasp the competence diversity needed in the public sector. Although this framework is a valuable approach, it is one of the very few attempts to systematically cluster competences and derive explicit categories. Further systematic research on which competences public sector employees need for the digitalization, has not been put on the IS research agenda.

3 Methodology

In order to uncover the status quo of research on competences in the public sector, we conducted a structured literature review, following the guidelines by [70]. Following their approach, we did not focus on one specific discipline, but widened the scope of our research. Contrary to the propositions by [70], we did not focus exclusively on journals but intentionally decided to include different kinds of publications, because of the fact that research on competences just recently gained in academic interest in the public sector [11], [12]. We first defined a set of relevant search terms that included synonyms to *competence* and *public administration* as well as *eGovernment*. We included the term e-government and variations of it, because a lot of research around the digitalization of public administrations is summarized under this umbrella term. In addition, we ran a second search in the database, this time reducing the search to articles also including keywords on digitalization, innovation or technologies. We used the database *SCOPUS* as it allows for various settings and includes a wide range of scientific outlets. The results were limited to outlets from the year 2000 onwards, because of the rapid changes technologies have undergone since then and it can be assumed that the needed competences have changed as well. As the study does not only aim at describing the status quo of research on competences in public administration but also on deriving an agenda for further research, this limitation is valid. Combining both searches led to a total set of 1235 articles. We excluded all results of doubtable quality, such as students' theses and grey literature as well as all non-scientific articles. Also, we only considered English articles, assuming that important research is published in English to make it accessible to a broader audience. We did, however, consider full as well as research in progress-papers and journal articles as well as conference proceedings and book chapters. Based on the title and abstract, we assessed the remaining articles with regard to their fit for this study's aim. The inclusion criterion was whether the development of *competences* in or for *public administrations* was a key topic of the articles. Although we rather included than excluded articles, also in borderline or unclear cases, in the end we only considered 63 articles to be relevant. We read these articles to evaluate whether they dealt with competences as a core topic or not. After a thorough pre-evaluation, a final set of 21 articles remained. 42 articles were not considered in the end, because they either purely focused on IT skills or they had a very specific target group that would not allow for comparability. We eliminated the majority of articles, because they only mentioned competences and/or skills on a side note, without specifying their nature and making them an explicit unit of analysis or the studies mentioned skills and competences as one factor influencing adoption decisions without further distinguishing the needed competences. With the remaining papers, we conducted a *forward* and *backward* search, applying the same inclusion and exclusion criteria as set forth above. All articles referencing one of the 21 articles were assessed which led to the inclusion of another 7 articles. For this search, we used *Google Scholar* as it yielded considerably more citations than SCOPUS. In the same way, we also examined all articles cited in the first 21 papers and included another 14 articles as relevant in our final data set.

4 Results

The final sample contains 42 articles (see Table 1). Out of these, 26 were published in journals, 13 of them were conference proceedings and 3 articles were book chapters. All articles were published between 2000 and 2018. Table 2 contains all identified sources, their shortened title, and outlet form. We also evaluated the extent to which the analyzed articles referred to each other as indicated by the columns 'citing' and 'cited by'. This analysis shows that only few articles refer to others in the sample and are frequently referred to by others such as [14], [21], [28] and [29].
For the analysis, we intentionally decided against following the chosen categorization of competences by the authors, although most articles already included some kind of categorization. We decided for breaking them up to identify patterns or differences ourselves. We did this to first eliminate possible ambiguities that could exist due to the authors' different foci and chosen terms and second to not overlook any skills that we would not have expected in certain categories. The articles' categorizations were coined very differently, given the difference in focus of the respective article. [62], for example, deal with skill sets of a successful collaborator, which is why their article naturally concentrates primarily on personality traits and soft skills. We found that similar skills were categorized differently in different articles; therefore, we summarized them in a new category.

Table 1. Articles identified by the literature review

Authors	*Outlet*	*Source*	*citing*	*cited by*
Leitner 2006 [29]	Cnf	D	-	[21], [30], [31], [32]
Banerjee et al., 2015 [33]	Jnl	D	-	-
McQuiston & Manoharan, 2017 [34]	Jnl	D	[35]	-
Michelucci et al., 2016 [36]	Jnl	D	[37]	-
Marzullo & Souza, 2011 [38]	Jnl	D	-	-
Gharawi et al., 2014 [39]	Cnf	D	[12], [20], [21], [40]	[41]
Hoefer, 2003 [42]	Jnl	D	-	-
Getha-Taylor & Lee, 2008 [43]	Jnl	D	-	-
Mancebo Fernandez et al., 2008 [44]	Jnl	D	-	-

Auluck & Levin, 2009 [45]	BC	D	-	-
Awortwi, 2010 [46]	Jnl	D	-	[47]
Haq, 2011 [48]	Jnl	D	-	-
Gupta et al., 2017 [28]	Jnl	D	[49], [50], [51], [52], [46]	-
Mincu, 2017 [53]	Jnl	D	[54]	-
Hunnius et al., 2015 [14]	Cnf	D	[55], [29], [21]	-
Yuryeva et al., 2015 [56]	Jnl	D	-	-
Iwasaki, 2014 [57]	Cnf	D	-	-
Gunn et al., 2014 [58]	Jnl	D	-	[59]
Hunnius & Schuppan, 2013 [21]	Cnf	D	[29]	[60], [61], [39]
o'Leary et al., 2012 [62]	Jnl	D	[63]	-
Schuppan, 2014 [64]	Cnf	D	[65]	[14]
Williams, 2002 [65]	Jnl	B	-	[62]
Thudugala & Weerawarana, 2013 [55]	Jnl	B	-	[14]
Virtanen, 2000 [52]	Jnl	B	-	[28]
Hondeghem & Vandermeulen, 2000 [66]	Jnl	B		[49]
Noordegraaf, 2000 [37]	Jnl	B	-	[40]
Mincu, 2016 [54]	Cnf	B	-	[53]
Marcovecchio et al., 2013 [40]	Cnf	B	-	-
Janowski et al., 2012 [12]	Cnf	B	[29], [30]	[39]
Schuppan, 2010 [30]	BC	B	[29]	[14]
Getha-Taylor & Morse, 2013 [51]	Jnl	B	-	[51]
Getha-Taylor, 2008 [63]	Jnl	B	-	[62]
Dawes, 2004 [35]	Jnl	B	-	[34]
Brans & Hondeghem, 2005 [50]	Jnl	B	[49], [66]	[28]
Bhatta, 2001 [49]	Jnl	B	[37], [52], [66]	[28]
Stare & Klun, 2018 [47]	Jnl	F	[46]	-
Ylinen & Pekkola, 2018 [41]	Cnf	F	[39]	-
Darling & Cunningham, 2016 [59]	Jnl	F	[58]	-
Ogonek et al., 2016 [60]	Cnf	F	[21]	-
Reichard & van der Krogt, 2014 [61]	Cnf	F	[21]	-
Schulz & Schuppan, 2011 [31]	BC	F	[29]	-
Schulz & Schuppan, 2012 [32]	Cnf	F	[29]	-

Key: Jnl=journal article; Cnf=conference paper; BC=book chapter; D=database search; B=backward search; F=forward search

In total, we identified two overarching categories, i.e. competences and personality traits. Those were split again into 12 sub-categories, identifying the concrete skills. Table 2 exhibits all sub-categories with exemplary mentions and corresponding references. The first identified category contains competences, which can be subsumed as all the skills and abilities that are related to a certain professional area or field, i.e. functional competences. The analysis of those competences yielded eight sub-competence areas. *Business skills* are mentioned most often. 33 of the 40 articles relate to some kind of business expertise. Competences in this sub-category relate to the organization and management of a public agency but are not necessarily domain-specific. Exemplary competences are project, program, performance and strategic management, which are among the most often cited ones. HR, economic, finance and accounting and marketing skills also fall into this category. Besides those business-related skills, *IS/IT skills* are the second most often cited competences that appear in 28 of the 40 articles. Again, skills listed in this category are not necessarily domain-specific and could be applied to competence frameworks for private organizations as well. Whereas some authors describe this category very generic with “IT content” [34] or “Familiarity with ICT systems” [33], other authors elaborate more detailed on those skills, like [38], [39] or [14], who mention concrete competence areas such as architecture, cyber security and managing information systems.

The next category *organization* unites abilities that are centered in the organization itself and, thus, are domain-specific. Competences in this sub-category deal with the characteristics of an organization, thus including skills such as administrative processes, organizational design or the identification with one’s agency. This sub-category is mentioned in 23 articles and, thus, seems to be an important asset in the job profile of the public sector. *(Public) Policy* was mentioned in 21 articles and includes skills such as knowledge about an agency’s policy area, policy planning and politics.
The category *law* occurred in 14 articles and is concerned with all legal aspects that arise in the work of a public administration. Those two sub-competence categories seem to be very specific to the public servant’s profile. The next sub-category summarizes competences that were either rarely mentioned or could not be grouped in a meaningful way to any other competence category, but still we deemed them worth mentioning. They are grouped under the name *other* and include competences in research (2 sources), socio-technical skills (5 sources) and professional experience (6 sources). After having categorized all the “hard”-knowledge based competences (functional competences), the last competence category lists all the *soft skills*, i.e. abilities that can be trained but do not count as knowledge. These competences are unrelated to any specific domain and include, for example, communication skills, teamwork, leadership, customer and service orientation and alike. It is noteworthy that those skills appear in 37 of the 40 articles and can thus be considered as highly important. Only in the categorization by [14] they are not explicitly mentioned. Within this group, especially the soft skill “leadership” is prominently mentioned in 25 articles. The second identified category are personality traits. Those traits are inherent to a personality and are not related to a specific (job-related) task. Furthermore, they cannot be acquired through formal education. We identified four sub-categories here. The most frequently mentioned one are *analytical skills*, which appeared in 24 articles. Analytical skills are more geared towards how a person tackles tasks and challenges.

Table 2. Identified competences and personality traits in identified articles

Type	***Area***	***exemplary competences/personality traits***
competence	business	strategic planning [29], [34] project management [33], [65]; program management [41], [57]; contact management [31], [41]; finance [36]; economics [47]; accounting [36],[42]
	IS/IT	management (information) systems [42], IT skills [14]; information systems [43], [47]; cyber security [39]; enterprise architecture [38]; technology management & assessment [57]
	organi-zation	organizational design [12], [14]; administrative processes & workflows [21],[37]; coordination/implementation [32], [48]; identification with agency [42], organizational theory [34]
	(public) policy	public policy [28],[46]; knowledge of agency's policy area [42]; social policy [42]; policy planning [45]; politics & political processes [65]; policy processes [31]
	law	administrative law [21]; legal aspects for data management [36]; legal tools [36]; legal aspects [39]; legal framework [40]; regulatory theory [53]
	other	professional experience [36], [42], [50]; evaluation & research [42], [61]; socio-technical skills [14], [35], [40]
	soft skills	leadership [56]; conflict management/negotiation [55]; (cross-cultural/unit, oral & written) communication [62]; mediation [62]; assertiveness [44]; influencing [54]; relationship [63];
personality	character traits	tolerance [33]; continuous learning [39]; creativity [48], commitment [45], [52]; tenacity & perseverance [44], flexibility [50]
	analytical	critical thinking [58]; analytical thinking[55]; strategic & innovative thinking [51]; decision making [49]; problem-solving [48], abstraction [41]
	self-management	self-organization [64]; self-control [28], self-awareness [62]; self-confidence [63]; self-reflection [64]
	other	psycho-social stability [45]

This category subsumes employee qualities indicating a view behind one's own horizon, i.e. critically reflecting on past events but also planning and envisioning scenarios, which requires thinking out of the box. The next sub-category consists of *character traits*. In contrast to the analytical skills, the articles here mostly describe more than one character trait and mention certain types of personality that public agencies need or look for. Character traits named here are, for example, tact, respect, patience, tolerance, and alike. Those seem to be preferable character traits in public administrations. This personality trait appeared in 23 of the 40 articles. The third sub-category of personality traits, mentioned in 14 articles is some kind of *self-management*. Again, competences in this category are more related to an employee's professional behavior. Those characteristics include self-organization, self-control and self-reflection, amongst other similar traits. The last sub-category under personality is coined *other*, which includes psycho-social stability. This sub-category was deemed so special, although it only appeared in one sole article that we decided to add it.

5 Discussion

The purpose of this research was to identify which competences public sector employees need from a research perspective. The analysis of articles allows drawing several conclusions, which we present in the following.

First, the literature review shows that only very few articles deal with the development of competences in the public sector. Moreover, only seven articles were published after or at the same time as the study by [14], which was a review of prior research on competences in the public sector as well. Although organizations such as the EU have recognized the need for the development and education of administration specific competences [27], this area has so far received only little scientific interest. While eGovernment scholars often mention competences as an important driver of the digitalization of the public sector, the study of competences in the public sector is still in its infancy. Although only few articles deal with the training of competences in the public sector, these studies analyzed a variety of different competences that public administrations need. Moreover, the studies do not differentiate whether the competences are needed everywhere to the same degree or not and if one employee should be trained equally in every competence category. Although. on a higher level, they differentiate between leaders in the public administration, e.g. [62] and employees, e.g. [44] and between IT related jobs, e.g. [38] and non-IT related jobs, e.g. [58], finer grained differentiations are not applied. For example, it remains unclear, whether the identified competences are needed in the public administration in general or whether they are task-specific. For example, an accounting clerk may need different competences than does the caseworker. Furthermore, ambiguities exist with regard to the relevance of specific competences for the different administrative levels. While some studies focus on a specific administrative level, e.g. [58], most of the articles do not provide information on whether the need for particular competences varies between the administrative levels. Again, the competences required of a municipal employee may differ from those required of an employee at the national level.

Second, our results show that all analyzed articles are empirical studies, whereas conceptual and theoretical approaches are missing completely. It is noteworthy that only few articles refer explicitly to conceptual works, e.g. [62], [34]. Instead, most articles directly introduce their method, e.g. [14], [56], rather than reviewing theoretical approaches related to the development of competences for the public sector, e.g. [36]. The focus of current research on empirical studies leads to the production of practically relevant knowledge but contributes to a lesser degree to the scientific body of knowledge. [34] point out: “As a *pracademic* field, public administration is based on the integration of academic concepts with its practical applications, as its boundary and scope are increasingly expanded in an ever-complex world”. However, the exclusive emphasis on empirical studies revealed by means of the literature review raises doubts if scholars in the field of competence research actually “[…] strive to maintain this balance” [34] between practice and academia. Accordingly, we find that only very few of the articles include references to another paper of our sample, although all articles deal with the development of eGovernment competences. This finding highlights a need for a better linkage of existing research. This fact is all the more astonishing as most of the papers are published in related outlets such as the proceedings of the *Hawaii International Conference on System Sciences* [12], [14], [21], [29], [64], *Public Administration* [50], [65], *Public Administration Review* [62], *International Review of Administrative Sciences* [46], *International Journal of Public Sector Management* [37], [52] or the *International Journal of E-Government Research* [38]. Thus, we propose to better link existing research on eGovernment competences and to conduct more conceptual research to establish a common theoretical understanding.

Third, the review shows that number of specific competences is very high. The proposed framework of competences for the public sector should, thus, be refined and validated – theoretically and empirically. As outlined above, the analyzed articles do not provide any information on whether the identified competences are task-specific or should be trained in general. Thus, the question remains to what degree employees in different departments should be educated in every category. To the best of our knowledge, only one study exists, which was implemented for the German IT planning council, that takes into consideration that competences may be task-specific and that the distribution of competences may depend on the specific job profile of an employee [67]. The authors distinguish the development of competences with regard to two aspects: First, public administration staff may need different competences according to their task or role. Second, the employees needing the same competences may not need them to the same degree. The depth of the competence acquaintance ranges from designing (highest level) over usage to knowledge (lowest level) and is based on the well-known taxonomy of educational objectives by [68]. In accordance with this study, we propose to not only categorize competences that are needed in digital public administrations overall but to differentiate the competences on two levels. On the *horizontal level*, competences are distinguished in relation to tasks within the public administration. For example, leaders of public agencies may need more business skills than does a case worker who may need more personal skills. Not every task may require every competence listed in the framework (see Table 2). On the *vertical level*, competences are distinguished according to the degree to which they have to be appropriated by the employees, i.e. although employees may need the same competence, e.g. competences in enterprise architecture, one may only need to know what this term comprises, whereas another employee may need to design such systems.

Fourth, the literature review revealed that business skills are far more often required than technical or other competences. Although the ubiquitous digitalization is reality, research seemingly puts less emphasis on the training and recruitment of technical competences. Moreover, it is noticeable that public administration-specific competences are less often required as compared to business skills. The analyzed articles are less context-specific as one might expect. About half of the identified competences were not domain or task-specific but more related to the employee's personality (e.g. soft skills, character traits etc.). Thus, we conclude that competences independent from the tasks are equally important for the public administration as are task-related competences. In relation to the overall competence framework, this leads us to the assumption that instead of looking for specialists, public administrations rather seek all-round talents. Given that researchers have recognized the necessity of the interplay between ICT and institutional settings as success factors for eGovernment implementation, e.g. [69], further research is needed that sheds light onto the specific competences public administrations need to successfully implement eGovernment, i.e. the framework proposed here needs to be validated. It also offers an estimation with regard to whether weighting of the competences as revealed by the literature review adequately addresses the challenges public administrations have to face in the digital age. Regarding the development of research on competences over time, it seems that the topic was of special interest in the beginning of the eGovernment area. 7 articles were published between 2000 and 2014. From 2005 to 2009, research on competences declined to 6 articles. It was not before the year 2010 that eGovernment competences received more attention with 17 articles published until 2014, which seems to go on, because since 2015 until today we could identify 12 articles.

6 Conclusion and Outlook

This study set out to identify the competences public sector employees need from a research perspective. To answer the research question, we conducted a structured literature review on competences in the public sector literature. Our review shows that to date only few studies exist that are concerned with competences in the public sector. Within these studies, we identified two overarching categories, competences and personality traits that could be further split up into 8 and 4 sub-categories respectively. We set up a research agenda to list all future research directions that can be derived from this study. It shows that besides the need for functional or task-related competences like organizational or management skills, there is also an increased need for soft skills and personality traits. Task independent competences seem to be equally important for the public domain as are functional competences, thus requiring public servants to be all-rounders rather than specialists in one specific area. Having said this, future research should also consider examining the respective competence depth more critically, since not every employee is required to embody one specific competence in the same way. Given the limited amount of studies identified in total and the finding that those are primarily of empirical nature, more research on this topic is needed in general and especially with regard to theoretical contributions that provide a solid basis as a commonly agreeable set for further research in this area to build on. After all, a diversified look into interdisciplinary competences, required by public servants working in a digital environment today, only picked up real speed in terms of scientific contributions by the year 2010. Thus, this topic has gained considerably in importance and more contributions are to be expected. The main limitation of this study is the choice of database that naturally fails to provide a complete picture of competence research in eGovernment. Thus, further research should look into other databases to round out and verify these results. Due to the pace of changes, induced by technology, competences also might undergo much more frequent changes than it was the case before. This is why more and particularly constant research is required to monitor these developments to competently address possibly new challenges.

References

1. World Bank: World Development Report 2016: Digital Dividends. Washington DC (2016)
2. Kim, G., Shin, B., Kim, K.K., Lee, H.G.: IT capabilities, process-oriented dynamic capabilities, and firm financial performance. J. Assoc. Inf. Syst. 12, 487 (2011)
3. Hoberg, P., Krcmar, H., Oswald, G., Welz, B.: Skills for digital transformation. IDT survey (2017)
4. European Commission: eGovernment Benchmark 2016. A turning point for eGovernment development in Europe? Luxembourg (2016)
5. Gascó, M.: New technologies and institutional change in public administration. Soc. Sci. Comput. Rev. 21, 6–14 (2003)
6. Reddick, C., Anthopoulos, L.: Information and communication technologies in public administration: innovations from developed countries. CRC Press, Boca Raton (2015)
7. Böhmann, T., Drews, P., Meyer-Blankart, C.: Digitale Exzellenz: Eine Bestandsaufnahme zur Digitalisierung deutscher Unternehmen und Behörden. Research Report. University of Hamburg (2015)
8. Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N., Ahlemann, F.: Digitalization: opportunity and challenge for the business and information systems engineering community. Bus. Inf. Sys. Eng. 59, 301–308 (2017)
9. Scholl, H.J.: Electronic government research: topical directions and preferences. In: Wimmer, M.A., Janssen, M., Scholl, H.J. (eds.) Electronic Government. 12th IFIP WG8.5 International Conference, EGOV 2013 Proceedings, Koblenz. pp. 1–13. Springer, Heidelberg (2013)
10. Scholl, H.J.: The EGOV Research Community: An Update on Where We Stand. In: Janssen, M., Scholl, H.J., Wimmer, M.A., Bannister, F. (eds.) Electronic Government. 13th IFIP WG 8.5 International Conference, EGOV 2014, Dublin, Proceedings. pp. 1–16. Springer, Heidelberg (2014)
11. Ogonek, N.: The Tale of e-Government: A Review of the Stories that Have Been Told So Far and What is Yet to Come. Proceedings of the 50th Hawaii International Conference on System Sciences, 2468–2477 (2017)
12. Janowski, T., Estevez, E., Adegboyega, O.: Conceptualizing electronic governance education. Proceedings of the 45th Hawaii International Conference on System Sciences, 2269–2278 (2012)
13. Janowski, T., Dellary, W., Davies, J.: Introduction to Electronic Government Education, Training and Professionalization Minitrack. Proceedings of the 46th Hawaii International Conference on System Sciences, 1662–1663 (2013)
14. Hunnius, S., Paulowitsch, B., Schuppan, T.: Does E-Government education meet competency requirements? An analysis of the German university system from international perspective. Proceedings of the 48th Hawaii International Conference on System Sciences, 2116–2123 (2015)
15. Nordhaug, O.: Human capital in organizations: Competence, training, and learning. Scandinavian University Press, Oslo (1993)
16. Albury, D.: Fostering innovation in public services. Public money and management 25, 51–56 (2005)
17. Besson, P., Rowe, F.: Strategizing information systems-enabled organizational transformation: A transdisciplinary review and new directions. J. Strategic Inf. Syst. 21, 103–124 (2012)

18. Danziger, J.N., Andersen, K.V.: The impacts of information technology on public administration: an analysis of empirical research from the "golden age" of transformation. Int. J. Public Admin. 25, 591–627 (2002)
19. Pérez-Escoda, A., Fernández-Villavicencio, N.G.: Digital competence in use: from DigComp 1 to DigComp 2. Proceedings of the 4th International Conference on Technological Ecosystems for Enhancing Multiculturality, 619–624 (2016)
20. Estevez, E., Janowski, T.: Landscaping government chief information officer education. Proceedings of the 46th Hawaii International Conference on System Sciences, 1684–1693 (2013)
21. Hunnius, S., Schuppan, T.: Competency requirements for transformational e-government. Proceedings of the 46th Hawaii International Conference on System Sciences, 1664–1673 (2013)
22. Kaiser, S.: Qualification Requirements in e-Government: The Need for Information Systems in Public Administration Education. In: Traunmüller, R. (ed.) Electronic Government. Third International Conference, EGOV 2004, Proceedings, Zaragoza., pp. 464–467. Springer, Heidelberg (2004)
23. Eshet, Y.: Digital literacy: A conceptual framework for survival skills in the digital era. J Educ. Multimedia Hypermedia 13, 93–106 (2004)
24. Devaux, A., Grand-Clement, S., Manville, C., Belanger, J.: Digital learning’s role in enabling inclusive skills development for a connected world. Proceedings of EMOOCs 2017, 104–111 (2017)
25. European Commission: Measuring Digital Skills Across the EU. EU wide indicators of Digital Competence. Brussels (2014)
26. SFIA Foundation: The Skills Framework for the Information Age (SFIA), https://www.sfia-online.org/en/reference-guide (Accessed: 16.01.2019)
27. A common European framework for ICT Professionals in all industry sectors, http://www.ecompetences.eu/ (Accessed: 16.01.2019)
28. Gupta, V., Chopra, S., Kakani, R.K.: Leadership competencies for effective public administration: a study of Indian Administrative Service officers. J. Asian Pub. Policy 11, 98–120 (2017)
29. Leitner, E.: eGovernment: People and skills in Europe's administrations. Proceedings of the 39th Hawaii International Conference on System Sciences, 1–9 (2006)
30. Schuppan, T.: E-government competencies: looking beyond technology. Handbook of Public Information Systems, 3rd ed. Taylor & Francis, Boca Raton, 353–370 (2010)
31. Schulz, S., Schuppan, T.: A European Framework for e-Government Competences. In: Stracke, C.M. (ed.) Competence Modelling for Human Resources Development and European Policies. Bridging Business, Education, and Training, pp. 142–154. Duisburg, Essen (2011)
32. Schulz, S., Schuppan, T.: Development of a European Framework for e-Government Competences. In: Lucke, J.v., Geiger, C.P., Kaiser, S., et al. (eds.) Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur. Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) & Rechtsinformatik (FTRI) 2012, pp. 47–58. GI, Bonn (2012)
33. Banerjee, P.K., Ma, L.C.K., Shroff, R.H.: E-governance competence: a framework. Elect. Gov., Int. J. 11, 171–184 (2015)
34. McQuiston, J.M., Manoharan, A.P.: Developing E-Government coursework through the NASPAA competencies framework. Teach. Public Admin. 35, 173–189 (2017)
35. Dawes, S.: Training the IT-Savvy Public Manager: Priorities and Strategies for Public Management Education. J Public Aff. Educ. 10, 5–17 (2004)

36. Michelucci, F.V., Marco, A. de, Tanda, A.: Defining the Role of the Smart-City Manager: An Analysis of Responsibilities and Skills. J Urban Technol 23, 23–42 (2016)
37. Noordegraaf, M.: Professional sense-makers: managerial competencies amidst ambiguity. Int J Public Sect Manag 13, 319–332 (2000)
38. Marzullo, F.P., Souza, J.M. de: New directions for IT governance in the Brazilian government. In: Weerakkody, V. (ed.) Applied Technology Integration in Governmental Organizations: New E-Government Research, pp. 313–334. IGI Global, Hershey, New York (2011)
39. Gharawi, M., Estevez, E., Janowski, T.: Identifying government chief information officer education and training needs: the case of Saudi Arabia. Proceedings of the 15th Annual International Conference on Digital Government Research, 280–289 (2014)
40. Marcovecchio, I., Estevez, E., Fillottrani, P.: Government Chief Information Officer (GCIO) Ontology - A Tool to Formalize the GCIO Function. In: Janowski, T., Holm, J., Estevez, E. (eds.) Proceedings of the 7th ICEGOV '13, pp. 32–41. ACM Press, New York (2013)
41. Ylinen, M., Pekkola, S.: Looking for a five-legged sheep. Identifying enterprise architects' skills and competencies. In: Zuiderwijk, A., Hinnant, C.C. (eds.) dg.o 2018: Proceedings of the 19th Annual International Conference on Digital Government Research, 2018, Delft, pp. 1–8. ACM Press, New York (2018)
42. Hoefer, R.: Administrative skills and degrees: The "best place" debate rages on. Admin. Soc. Work 27, 25–46 (2003)
43. Getha-Taylor, H., Lee, J.: Changing competencies for human resource management: examining e-government scorecard and Federal Human Capital Survey evidence. Int. J. Public Sect. Perform. Manag. 1, 150–166 (2008)
44. Mancebo Fernandez, N., Rafart Serra, A., Bikfalvi, A., Munoz Soler, M.D.: Aligning leadership and competences in recruitment and staff development: an empirical analysis in the context of regional public administration. Int. J Public Sect. Perform. Manage. 1, 167–182 (2008)
45. Auluck, R., Levin, R.: Chapter III. Developing Gobal Public Servants for Good Governance and Sustainable Development. In: Pagaza, I.P., Argyriades, D. (eds.) Winning the Needed Change: Saving our Planet Earth, pp. 48–63. IOS Press, Amsterdam, Berlin, Oxford, Tokyo, Washington DC (2009)
46. Awortwi, N.: Building new competencies for government administrators and managers in an era of public sector reforms: the case of Mozambique. Int. Rev. Adm. Sci. 76, 723–748 (2010)
47. Stare, J., Klun, M.: Required competencies in public administration study programs. Transylv. Rev. Adm. Sci. 55, 80–97 (2018)
48. Haq, S.: Ethics and leadership skills in the public service. Procd. Soc. Behv. 15, 2792–2796 (2011)
49. Bhatta, G.: Enabling the Cream to Rise to the Top: A Cross-Jurisdictional Comparison of Competencies for Senior Managers in the Public Sector. Public Perform. Manag. 25, 194–207 (2001)
50. Brans, M., Hondeghem, A.: Competency Frameworks in The Belgian Governments: Causes, Construction and Contents. Public Admin. 83, 823–837 (2005)
51. Getha-Taylor, H., Morse, R.: Collaborative Leadership Development For Local Government Officials: Exploring Competencies and Program Impact. Public Admin. Quart. 37, 71–102 (2013)
52. Virtanen, T.: Changing competences of public managers: tensions in commitment. Int. J. Public Sect. Manag. 13, 333–341 (2000)

53. Mincu, C.: Sharpening Leadership Edge in the Public Sector and Private Business for Adaptive-to-Change Organizations. Qual. Access Success 18, 9–16 (2017)
54. Mincu, C.: Building Transformational Leadership Competencies in the Public Sector to Secure Quality Service. International Conference of the Institute for Business Administration in Bucharest, 2016, 29–33 (2016)
55. Thudugala, K., Weerawarana, S.M.: Requisite competencies for government Chief Information Officer in Sri Lanka. Global Bus. Econ. Res. J. 2, 1–11 (2013)
56. Yuryeva, O.V., Butov, G.N., Malganova, I.G., Pratchenko, O.V.: Professionalism of Civil Servants as the Factor of Public Administration Efficiency Growth. Mediterr. J. Soc. Sci. 6, 481 (2015)
57. Iwasaki, N.: The gap between CIO core competencies and the real roles of CIOs. Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, 244–250 (2014)
58. Gunn, E., Cunningham, J.B., MacGregor, J.N.: Developing benchmarks for guiding CAO performance. Local Gov. Stud. 40, 851–868 (2014)
59. Darling, S.D., Cunningham, J.B.: Underlying values and competencies of public and private sector managers. Asian Educ. Dev. Stud. 5, 371–387 (2016)
60. Ogonek, N., Gorbacheva, E., Räckers, M., Becker, J., Krimmer, R., Broucker, B., Crompvoets, J.: Towards Efficient EGovernment: Identifying Important Competencies for EGovernment in European Public Administrations. In: Scholl, H.J., Glassey, O., Janssen, M., et al. (eds.) Electronic Government and Electronic Participation. Proceedings of IFIP EGOV and ePart 2016. IOS Press, Amsterdam, Berlin, Washington DC (2016)
61. Reichard, C., van der Krogt, T.: Towards a set of specific competences for academic degree programmes in Public Administration in Europe. 22nd NISPAcee Conference in Budapest, 22.-24 May, 2014, 1–19 (2014)
62. O'Leary, R., Choi, Y., Gerard, C.M.: The skill set of the successful collaborator. Pub. Admin. Rev. 72, S70-S83 (2012)
63. Getha-Taylor, H.: Identifying Collaborative Competencies. Rev. Public Pers. Adm. 28, 103–119 (2008)
64. Schuppan, T.: E-Government at Work Level: Skilling or De-skilling? Proceedings of the 47th Hawaii International Conference on System Sciences, 1927–1934 (2014)
65. Williams, P.: The Competent Boundary Spanner. Public Admin. 80, 103–124 (2002)
66. Hondeghem, A., Vandermeulen, F.: Competency management in the Flemish and Dutch civil service. Int. J. Public Sect. Manag. 13, 342–353 (2000)
67. Becker, J., Greger, V., Heger, O., Jahn, K., Krcmar, H., Müller, H., Niehaves, B., Ogonek, N., Räckers, M., Schuppan, T.: E-Government-Kompetenz (2016)
68. Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R.: Taxonomy of educational objectives: The classification of educational goals. Handbook I: The Cognitive Domain. Longman, New York (1956)
69. Cordella, A.: E-Government Success: How to Account for ICT, Administrative Rationalization, and Institutional Change. In: Gil-Garcia, J.R. (ed.) E-Government Success Factors and Measures. Theories, Concepts, and Methodologies, pp. 40–51. IGI Global, Hershey, New York (2013)
70. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a Literature Review. MIS Quart. 26, xiii–xxiii (2002)

Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents

Sebastian Hobert[1], Raphael Meyer von Wolff[1]

[1] University of Goettingen, Goettingen, Germany
{shobert, r.meyervonwolff}@uni-goettingen.de

Abstract. In this paper, we present the current state of the art of using conversational agents for educational purposes. These so-called pedagogical conversational agents are a specialized type of e-learning and intelligent tutoring systems. The main difference to traditional e-learning and intelligent tutoring systems is that they interact with learners using natural language dialogs, e.g. in the form of chatbots. For the sake of our research project, we analyzed current trends in the research stream as well as research gaps. Our results show for instance that (1) there is a trend towards using mobile conversational agents in education, (2) a proper generalization of existing research results (e.g. design knowledge) is missing, and (3) there is a need for comprehensive in-depth evaluation studies and corresponding process models. Based on our results, we outline a research agenda for future research studies.

Keywords: pedagogical conversational agent, intelligent tutoring system, technology-enhance learning, chatbot, natural language processing.

1 Introduction

Intelligent learning systems can be used in educational contexts to improve learning processes [1]. In traditional learning settings (like in university lectures or in seminars in vocational trainings) it is challenging to provide individualized learning support to learners or to respond to every personal demand of learners individually. This challenge is targeted by intelligent tutoring systems [2]. They promise to provide individualized and personalized learning support regardless of the number of learners.

Due to the increasing spread and media attention of artificial systems as well as machine learning applications, the demand of learners and teachers for intelligent learning systems rises. The availability of intelligent dialog-based systems like Facebook Messenger bots increases the demands further. This results in a growing research interest *in pedagogical conversational agents* as part of the research stream of intelligent tutoring systems [3]. Research projects targeting pedagogical conversational agents deal with the question of how to provide proper learning support to learners via natural language interfaces. Thus, these research interests combine the pedagogical view on individualization of learning processes with promising technologies like

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

artificial intelligence and natural language processing to provide easy to use learning systems.

The use of natural language user interfaces for learning systems seems especially promising as learners already use messenger-like systems in private life commonly. According to recent representative surveys in Germany, over 90 % of those questioned use messenger services – over 75 % of them on a daily bases [4]. Thus, using chat-based systems including chatbots (i.e. conversational agents for non-educational use cases like FAQ bots [5]) are common. So, it is to expect that learners need (almost) no time to get used to natural language-based learning systems.

Even though the use of pedagogical conversational agents seems to be promising, the success of such systems should not be pledged early. In fact, researchers should carefully analyze whether the demand of users towards using the technology is just because of the current rise of artificial intelligence in general or whether it really enables improvements in learning processes. As a first step to consolidate the existing research interest and to provide a structured guidance towards a research agenda for design-oriented research, we outline the results of a structured literature review targeting text-based pedagogical conversational agents in this research paper. Thus, we ask the following research questions to describe our research targets:

RQ1: What is the current state of the art of using pedagogical conversational agents in education?

RQ2: Which design-oriented research gaps exist in the current research on pedagogical conversational agents?

To answer these research questions, the remainder of this paper is structured as follows: First, we briefly describe the term *pedagogical conversational agent* and outline related concepts like *intelligent tutoring systems*. Based on these basic terms, we describe the research approach of our literature review process in section 3. Following, we present our results focusing on four main perspectives (time, technical, didactical and methodical perspectives) in section 4 and discuss them in section 5 to outline a future research agenda. Finally, we briefly summarize the findings in the conclusion.

2 Basic Terms

Pedagogical conversational agents can be defined as a special form of learning applications that interact with learners individually [6]. The conversation of those agents usually takes place using natural language [2]. From a technical perspective, there exist two common types of pedagogical conversational agents:

1. *Messenger-like agents* that use common chat interfaces (e.g. known from WhatsApp or similar chat interfaces),
2. *Embodied conversational agents* (like game characters or avatars) that consist of a (virtual) representation of a person in virtual environments and communicate either via text-based or voice-based language.

Whereas in the past embodied agents were common, nowadays especially messenger-like agents (a.k.a. chatbots) are widespread. The reason is that messenger apps are considered to be easy to use, because “interaction takes place through messaging applications to which students are already very keen on” [7].

To enable a language-based communication, pedagogical conversational agents usually combine technical methods from natural language processing and machine learning. Thus, pedagogical conversation agents have the possibility to act in different human roles like tutors, students or colleagues [2]. Because of this stream in research, pedagogical conversational agents have similarities and intersections with intelligent tutoring systems [3, 8]. Both software systems are learning applications that aim at providing (individualized) assistance to learners [6, 8]. The combination of both definitions is also known as “conversational intelligent tutoring system” [3].

3 Literature Review Process

To answer the research questions raised in section 1, we conducted a systematic literature review. We followed established methodical approaches in performing literature analyses including those of *Cooper* [9], *Webster/Watson* [10] and *Fettke* [11]. Furthermore, we especially took the work of *vom Brocke et al.* [12, 13] into account that outlines recommendations on how to deal with the existing “literature overload” [12]. By adopting the *Framework for literature reviewing* [13], we apply the proposed five consecutive research steps that we briefly describe in the following. During the whole literature review process, we documented all process steps in a search protocol as proposed by [12]. To ensure the traceability of our research process, we attached a condensed version of the search protocol including our list of reviewing criteria at https://publikationen.as.wiwi.uni-goettingen.de/getfile?DateiID=739.

3.1 Definition of Review Scope

As we intend to identify the current state of the art of using pedagogical conversational agents for supporting learners especially in higher education as well as in workplace learning, we primarily focus our literature review on the research outcomes as well as on applications of published research papers. In this literature review, our goal is to identify central aspects on a conceptual level. As we intend to share the results not only with specialized scholars but also with interested practitioners, we synthesize the outcomes to highlight important results. Table 1 summarized the characteristics of the review by adapting the taxonomy of [9].

Table 1. Definition of review scope (adapted from [9])

Focus	**Outcomes**	Methods	Theories	**Applications**
Organization	Historical	**Conceptual**	Methodological	
Perspective	Neutral representation	**Espousal of position**		
Coverage	**Exhaustive** (in analyzed sources)	Representative	Central or Pivotal	
Audience	**Specialized scholar**	**General scholars**	**Practitioners**	General public

3.2 Conceptualization of the Topic

As a basis for the conceptualization of the topic of pedagogical conversational agents in the context of this paper, we rely on the definitions outlined in section 2. Using those definitions as a starting point as well as the works of [14, 15], we derive the following systematization that we use to classify the corpus.

First, we conceptualize the relevant papers based on technical considerations (see Table 2). In particular, we focus on the *type* of pedagogical conversational agents that are examined in the paper corpus: (1) Messenger-like conversational agents that provide user-interfaces in forms of text-based chats. (2) Embodied conversational agents that include visualizations of the agent like an animated virtual avatar. We also take the target *platform* of the considered conversational agents into account: (1) Mobile-first agents that focus mainly on smartphones (and tablet computers). (2) Web-based agents that are usable on any platform but do not follow a mobile-first approach. (3) Others, like standalone applications, that are limited to specific desktop operating systems.

Table 2. Criteria for conceptualizing the literature review (1/2)

<table>
<tr><td>Type</td><td colspan="2">Messenger-like conversational agent</td><td colspan="2">Embodied conversational agent</td></tr>
<tr><td>Platform</td><td>Mobile-first</td><td colspan="2">Web-based</td><td>Standalone/Others</td></tr>
</table>

In addition to the classification based on technical aspects, we distinguish the analyzed papers by the targeted *learning settings*. In this category, we differ between formal learning settings (like using a pedagogical conversational agent at a university during seminar sessions) and non-formal learning settings (like using an agent at home for self-study). Even though there exist more sophisticated forms of distinguishing learning settings, this seems sufficient for the given view on the literature. Furthermore, we categorize the articles by *learning form* and distinguish the following types: isolated, collective, situated and collaborative learning (see [14, 15] for detailed explanations of the learning forms). Finally, we consider the *content* view of the papers in which we differ single-topic learning content or multi-topic learning content. We distinguish both because pedagogical conversational agents (1) can be built for just a single purpose (e.g. training a special situation in language learning; single-topic) or (2) can be used to assist learners during a series of lectures or seminars (multi-topic). Table 3 summarizes the conceptualization that is a basis for the conduction of the literature review.

Table 3. Criteria for conceptualizing the literature review (2/2) based on [14, 15]

<table>
<tr><td>Learning setting</td><td colspan="2">Formal learning settings (e.g. at a university while attending a seminar)</td><td colspan="2">Non-formal learning settings (e.g. self-study at home)</td></tr>
<tr><td>Learning form</td><td>Isolated learning</td><td>Collective learning</td><td>Situated learning</td><td>Collaborative learning</td></tr>
<tr><td>Content</td><td colspan="2">Single-topic learning content</td><td colspan="2">Multiple-topic learning content</td></tr>
</table>

3.3 Literature Search

To identify relevant publications according to the review scope (see Table 1 above), we conducted a keyword search in bibliographic databases in June 2018. As the focus of the literature review can be assigned to the field of technology-enhanced learning, we selected databases that have at least partly a technology-oriented view. We further broadened the search scope by including all eight journals of the AIS Senior Scholars' Basket of Journals [16]. In doing so, we included in the following sources:

Table 4. Selection of searched sources

Journals	Scientific databases
– European Journal of Information Systems – Information Systems Journal – Information Systems Research – Journal of AIS – Journal of Information Technology – Journal of MIS – Journal of Strategic Information Systems – MIS Quarterly	– AIS Electronic Library – ACM Digital Library – IEEE Xplore Digitial Library – ScienceDirect – EBSCOhost Business Source Complete

To perform the literature search, we used the search terms listed in Table 5 in all selected databases and journal sources. In doing so, we included articles that match our research target directly (search term #1) as well as articles that cover most likely closely related topics (search terms #2 and #3). In total, we obtained approx. 550 papers.

To evaluate the search results for relevance, we defined criteria for inclusion as well as exclusion in accordance with [12]. Using those criteria, we reviewed titles and abstracts of all search results in a first step. Based on the resulting corpus, we reviewed the full text to the best of our knowledge and came up with a total number of 41 papers that represent original findings as the final corpus.

Table 5. Overview over search terms

Search terms	
#1	"pedagogical conversational agent"
#2	"smart teaching assistant" OR "AI teaching assistant" OR "artificial intelligence teaching assistant" OR "virtual teaching assistant"
#3	(chatbot OR chatterbot OR talkbot OR "interactive agent" OR "dialog system" OR "conversational agent") AND (learning OR teaching)

3.4 Literature Analysis and Synthesis and Proposition of Research Agenda

In the next two phases of the literature review process, we first analyze and synthesize the final corpus of relevant articles. We first examine the corpus using a time perspective before we analyzed its content in full depth by focusing on technical, didactical and methodical aspects (see section 4). Second, we discuss trends and research gaps in the corpus in order to derive a research agenda that is shown in section 5.

4 Results

In the following, we outline the results of our literature review using four perspectives. First, we focus on the time perspective of the paper corpus. Afterward, we examine the technical foundation, before we take a didactical perspective. Finally, we switch to a methodical perspective and outline how the researchers of the identified relevant papers conducted their research.

4.1 Time Perspective

A first descriptive analysis of the corpus regarding the distribution of publications per year shows an increasing interest in recent years. In 2017, the number of relevant publications reached an all-time high in the considered timeframe. Furthermore, the number of papers published in the first month of 2018 already reached the second highest number of publications. After interpolating the publications for 2018 based on the first six months (see Figure 1), we expect that the number of relevant publications in 2018 will exceed the current high. Thus, we conclude that there is an increasing interest apparent.

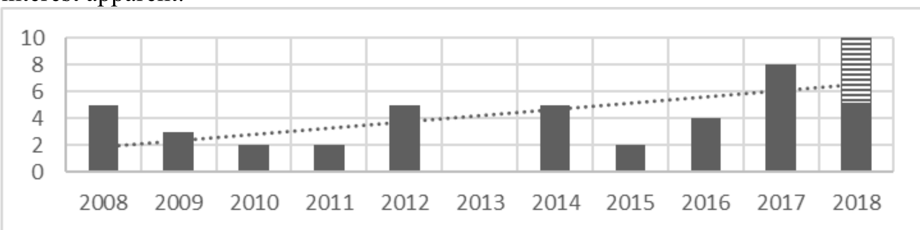


Figure 1. Distribution per year (grey shaded: 2018 interpolated based on June 2018)

4.2 Technical Perspective

As the first step of the content analysis, we categorized our final paper corpus by looking at the different technical types of agents as well as on their target platforms.

Agent Type. First, the distribution outlined in Figure 2 shows that the number of messenger-like conversational agents is more than twice the number of articles targeting at embodied conversational agents.

Figure 2. Distribution of articles per conversational agent type

Furthermore, the average publication date of all articles targeting messenger-like conversational agents (MCA: 2013.2) is approx. one year newer compared to the average of all embodied conversational agent papers (ECA: 2012.2). Thus, there is a trend present to use messenger-like conversational agents instead of embodied conversational agents. Even though the difference seems not to be huge, the trends become apparent when looking at the time lines and trend lines in Figure 3.

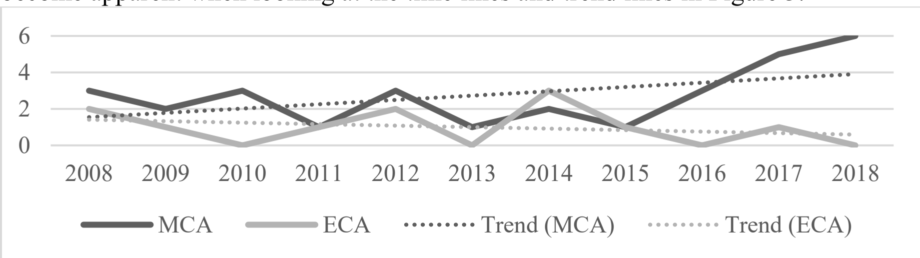


Figure 3. Trends per conversational agent type (year 2018 interpolated based on June 2018)

Target Platform. By analyzing the technical foundations of the identified conversational agents independently of the type, we could identify a trend towards implementing conversational agents for mobile platforms. Whereas mobile conversational agents are targeted in the paper corpus since 2016 with an increasing amount, the number of web-based or standalone conversational agents is decreasing. This is also reflected in the average publication date: Mobile agents have been published in average in 2017. In contrast to that, papers in both other categories have an average publication date of approx. 2012. However, it needs to be remarked that some web-based conversational agents are implemented responsively (e.g. [17]), i.e. they can be used on smartphones as well. However, if the primary focus of such prototypes is a desktop computer, we did not count them as mobile-first approaches.

Technical system architectures. Focusing on the technical system architectures of pedagogical conversational agents, we could identify that there is a large variety of system architectures present. Nevertheless, most implementations are based on a client-server architecture where natural language processing is done on the server side (e.g. [18]). However, the language processing steps vary in the different implementations from simple command-based matching approaches to the application of advanced machine learning based toolkits. Concerning the storage of data, some authors store, for instance, predefined question and answer tuples in relational databases, whereas others use AIML-files to store patterns and related answering templates (e.g. [19]). In some cases, answers are based on learning objects or learning paths (e.g. [20]). On the client side, user interfaces differ as well: Some agents are implemented as standalone clients (e.g. [21]) whereas others are integrated into third-party messaging platforms (e.g. [7]) or learning systems (e.g. [19]).

Summary. We conclude that the current research mainly focuses on messenger-like interfaces in mobile settings. On the one hand, the trend towards researching messenger-like agents seems like a reduction of complexity as the user interfaces get simpler. However, on the other hand, we could not identify a reduction in complexity in the system design as we could not identify a uniform architecture approach.

4.3 Didactical Perspective

Whereas the technical perspective focuses on the underlying software of conversational agents, we use the didactical perspective to analyze the educational application scenarios and the learning settings of pedagogical conversational agents.

Learning Setting. First, we categorized the papers of the corpus into formal or non-formal learning settings. As a result, we can conclude that only approx. 25 % of the papers that we could categorize[1] focus on formal learning settings. An exemplary paper is [22] in which the authors developed a virtual embodied avatar that can be used to simulate a virtual patient. Medical students can interview the embodied avatar to train diagnosing the avatar's diseases. The learning situation can be categorized as formal learning setting as the virtual embodied avatar is used by the students in a laboratory and not in a non-formal situation (e.g. at home). Another example of a mostly formal setting is presented by [23]. In this case, a pedagogical agent is used in an online tutoring task to guide the learners. As the learners need to communicate with the agents as a homework task, we categorized it as a formal learning setting.

Even though some papers cover formal learning settings, according to our analysis most pedagogical conversational agents target non-formal settings and provide communication possibilities to learners independently of a specific location, time or learning environment (e.g. course or lecture). However, in many cases the learning setting depends on the concrete use: In many cases, learners can interact with the agent location independent whenever they want using a smartphone or a desktop computer. But if the use of the pedagogical agent is integrated into the curriculum (like homework; see e.g. [23]), the transition to formal settings is smooth. Selected examples of non-formal learning settings are the prototypes *Oskar* [8] that can be used by learners for training the use of the database language SQL via an intelligent tutoring system and *Charlie* [19], which is a natural language user interface to the *INtelligent Educational System (INES)*.

Learning Form. The large number of papers covering non-formal learning settings is also reflected in the learning forms. As many conversational agents can be used by learners in any location independent of the user's environment, the learning often (approx. 66 %) takes place in an isolated way (i.e. learners interact with the conversational agent without any interaction with other learners, human tutors or lecturers). The remaining learning forms (collective, situated and collaborated learning) only take place in approx. 14 %, 3 %, and 17 % respectively.

Topic Focus. According to our statistics, the pedagogical conversational agents that have been researched in the paper corpus are almost equally distributed among single-topic agents or multi-topic agents. In many cases, the authors of the papers state that the agents are intended to support a specific learning scenario (e.g. the agent *Dr. Roland* [24] focuses on supporting learners to solve math problems; [25] is able to ask very specific questions and give hints about a special simulation). In other cases, the described agents can be used in almost any learning scenario, because teachers or lecturers are able to edit the learning content or add additional content via control

[1] In some cases, we could not categorize the papers, because too few information was available.

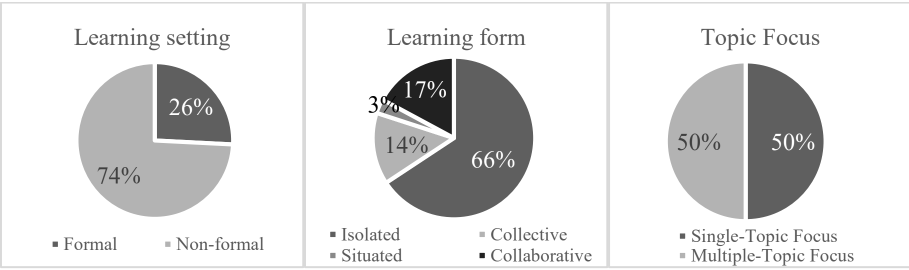


Figure 4. Summary of the didactical perspective on the paper corpus

panels. Exemplary systems that we identified are *MentorChat* (see e.g. [26]) or the agent by [27] that provides a so-called *Learning Objects Authoring Interfaces*.

Summary. Based on these results of taking an didactical perspective on pedagogical conversational agents (see Figure 4), we conclude that most available agents target non-formal learning situations in which learners interact with the agents alone (in an isolated way). However, it has not yet been conclusively investigated whether these use cases are beneficial over use cases in formal learning settings. Nevertheless, it needs to be remarked that this observation does not imply that such pedagogical conversational agents just support simple learning cases as known from simple mobile learning applications (like vocabulary training). On the contrary, pedagogical conversational agents are able to provide interactive, natural language-based opportunities to convey learning content in a way that was previously only possible in human-to-human training settings (like in classroom training or individual tutoring). Consequently, the isolated learning form as we used it in our classification only means that the learner uses the agent alone without any interaction with other learners. Nevertheless, the learners are not really isolated as they interact with an automated, virtual, but natural language-based chat partner.

We could not determine any trend regarding the topic focus of agents. On the one hand, agents that allow an administrator to configure and provide multiple topics seem useful. On the other hand, many researchers focus on single-topic agents that are specialized to fulfill a given learning task.

4.4 Methodical Perspective

In addition to the analysis of the content of the papers, we also examined the applied methods. Thus, we aim at identifying how researchers in the domain of pedagogical conversational agents are conducting their research.

First, we need to acknowledge that the methods that are described in the papers differ because the research field of pedagogical conversational agents is interdisciplinary: Researchers from the domains of computer science, information systems, pedagogy, and psychology are participating. This makes it difficult to classify the methods, as they differ depending on the discipline or the individual research background. For this reason, we have opted to an aggregated, qualitative view. Two aspects are particularly noteworthy: Prototype development and evaluations.

Prototype Development. In a majority of the papers, the authors based their research on implemented software prototypes (including *Charlie* [19], *Oskar* [8], *MentorChat (e.g. [26])*, *Dr. Roland* [24], *Ville* [28], *DEAL* [28], *AutoTutor* [29], *WrenchTalker* [30], *CiboPoli Bot* [31] and many unnamed more). From a methodical view, it is difficult to classify the methods as often there are not enough details available, but they seem to belong to the methods of design science research. However, to our knowledge, none of the research projects covers the whole design science research cycle starting with the problem identification phase and continuing with the definition of objectives, design, implementation, demonstration, evaluation and communication [32].

Evaluation. Focusing on the evaluation step, we identified two major directions: On the one hand multiple researchers chose Wizard-of-Oz experiments as an option to evaluate the potentials of pedagogical conversational agents in a simulated experiment. On the other hand, field studies were conducted to observe the operability, acceptance and beneficial value of the pedagogical conversational agents. However, there is no uniform evaluation method approach.

Summary. Based on the methodical view of the paper corpus, we conclude that the applied methods are quite heterogeneous. Only in the evaluation of pedagogical conversational agents, we can observe a trend towards using Wizard-of-Oz experiments as well as field studies. Considering the other steps of typical design science research cycles, we cannot identify a consolidation towards using specific methods. In many cases, only parts of the design science cycle are covered. Often the authors focus either on a technical view (covering conceptualization and implementation) or on a pedagogical view (covering learning scenario and evaluation). Comprehensive approaches are usually missing. In particular, it needs to be noted that neither generalizable requirements nor design theories are presented.

5 Discussion of Trends and Future Research Agenda

The first research goal (RQ1) of our study was to identify the current state of the art of using conversational agents for educational purposes. For this reason, we conducted a structured literature review and analyzed the resulting final paper corpus from four perspectives: From analyzing the time perspective, we determined that there is an increasing interest in research about pedagogical conversational agents. This becomes apparent as the total number of relevant papers reaches an all-time high in 2017 and we expect an even increasing number of publications by the end of 2018.

The technical perspective outlines that both, messenger-like conversational agents as well as embodied conversational agents, are objects of research. However, we can observe an upward trend in the number of publications focusing on messenger-like conversational agents whereas the number of embodied conversational agents seems to be decreasing. We explain this mainly with the increasing popularity of messenger apps in private life [4]. Additionally, an increasing number of messenger platforms and social networks started to provide APIs that can be used for developing chatbots (e.g. Slack, Facebook or Telegram). This resulted in a growing number of chatbot applications. In addition to that, an increasing interest of researchers and practitioners to adapt methods known from the machine learning or artificial intelligent domain can be observed (e.g., complex natural language understanding and generation algorithms). Due to this, intelligent chatbots receive more attention in research.

Our analysis of the didactical perspective shows that pedagogical conversational agents are most often designed for non-formal learning settings. Thus, learners can use the agents anywhere and anytime because of the common mobile or web-based implementation. For this reason, learning often takes place isolated from other learners but in interaction with a conversational agent. Regarding the topic focus, we could not identify a trend either towards single-topic or multiple-topic agents. However, we argue that the quite large number of single-topic agents is especially related to that fact that it is easier for researchers to limit the complexity of the software to a single topic, as this is often sufficient for conducting an evaluation. In some cases, we assume that the software artifacts are even capable of supporting multiple-topic scenarios, but the authors did not state this explicitly in the written papers. Thus, we argue that there is a trend towards multiple-topic agents when the focus is on actually using them in real settings and not only in laboratory experiments.

Finally, the methodical perspective shows the presence of a heterogeneous method mix. This represents the fact that the field of pedagogical conversational agents is quite interdisciplinary. As information system researchers, we must admit that complete design science research cycles and especially generalizable results in almost all design science research steps are missing.

Trending Characteristics. Based on our literature review and our interpretation of the results in the discussion above, we summarize the trends for pedagogical conversational agents in the following Table 6.

Table 6. Trending characteristics of pedagogical conversational agents

<table>
<tr><th>Type</th><td colspan="6">Messenger-like conversational agent</td><td colspan="6">Embodied conversational agent</td></tr>
<tr><th>Content</th><td colspan="4">Mobile-first</td><td colspan="4">Web-based</td><td colspan="4">Other</td></tr>
<tr><th>Learning setting</th><td colspan="6">Formal learning settings (e.g. at a university while attending a seminar)</td><td colspan="6">Non-formal learning settings (e.g. self-study)</td></tr>
<tr><th>Learning form</th><td colspan="3">Isolated learning</td><td colspan="3">Collective learning</td><td colspan="3">Situated learning</td><td colspan="3">Collaborative learning</td></tr>
<tr><th>Content</th><td colspan="6">Single-topic learning content</td><td colspan="6">Multiple-topic learning content</td></tr>
</table>

Future Research Agenda. With these results, we state the following research gaps and research opportunities:

(1) Need for generalized design knowledge. As most publications in the field focus on specific implementations of pedagogical conversational agents and miss to provide in-depth transferable insights, we propose that researchers should focus especially on the generalization of their design results. We are aware that it might be difficult to propose requirements that are generalized but also meaningful. Nevertheless, proposing generalizable system architectures for pedagogical conversational agents that can be transferred for different learning settings will be useful for researchers and

practitioners. As information system researchers, we propose to conduct multiple design science research cycles as a proper method to achieve this.

(2) Need for comprehensive in-depth evaluations. In the paper corpus, we identified many evaluations of pedagogical conversational agents. However, those papers often focus on very specific evaluation targets. Comprehensive evaluations covering multiple aspects (like learning success, technology acceptance, software quality, algorithmic quality, suitability of application scenarios) are missing. Even though an in-depth analysis of the evaluation methods was not in the focus of this article, we recognized that this needs to be addressed in the future. In particular, it would be meaningful to provide researchers with a detailed overview of suited evaluation methods. This is especially important for conducting comprehensive evaluations in research studies focusing on pedagogical conversational agents as the research field is interdisciplinary.

(3) Need for process models. Currently, the range of approaches on how to conduct research in the interdisciplinary field of pedagogical conversation agents is huge, because there is no uniform procedure. We assume that a common understanding of (a) how to develop and use pedagogical conversational agents in practice-oriented projects and (b) how to evaluate those agents comprehensively would advance the research field. For this reason, we propose that future research studies should focus on providing process models that cover both, design steps and evaluation methods. In particular, it would be helpful for design-oriented researchers as well as for practitioners to obtain a guideline that describes at which stage of the development process which evaluation methods are useful (e.g. conducting Wizard-of-Oz experiments seem especially useful in an early stage of the design process whereas field experiments are more useful after a functional prototype is available).

6 Conclusion

In this research study, we evaluated prior research papers in the interdisciplinary research stream of pedagogical conversational agents by taking time, technical, didactical and methodical perspectives on the literature base. Using this approach, we identified the state of the art and outlined trends that are present in research projects targeting at pedagogical conversational agents. Additionally, we proposed a research agenda. Possible limitations of this study lie in the selection and interpretation process of the analyzed papers. We are aware that these steps are dependent on the judgment of the individual researchers. Through a systematic literature analysis approach in which we defined inclusion and exclusion criteria, we tried to minimize the subjective influence as much as possible. Our results can contribute to both, research and practice: Researchers can base future projects on our research agenda to develop the field further. Additionally, our research might be helpful for practitioners. Especially developers of chatbot applications might use our results as a starting point to inform themselves about current trends in the field of pedagogical conversational agents.

References

1. Mikic Fonte, F.A., Burguillo, J.C., Nistal, M.L.: An intelligent tutoring module controlled by BDI agents for an e-learning platform. Expert Systems with Applications 39, 7546–7554 (2012)
2. Kerly, A., Hall, P., Bull, S.: Bringing chatbots into education: Towards natural language negotiation of open learner models. Knowledge-Based Systems 20, 177–185 (2007)
3. Crockett, K., Latham, A., Whitton, N.: On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees. International Journal of Human-Computer Studies 97, 98–115 (2017)
4. Schlobinski, P. and Siever, T.: Sprachliche Kommunikation in der digitalen Welt. Eine repräsentative Umfrage, durchgeführt von forsa, http://www.mediensprache.net/networx/networx-80.pdf (Accessed: 17.07.2018)
5. Meyer von Wolff, R., Hobert, S., Schumann, M.: How May I Help You? – State of the Art and Open Research Questions for Chatbots at the Digital Workplace. In: Proceedings of the 52th Hawaii International Conference on System Sciences. Maui, Hawaii (2019)
6. Tamayo-Moreno, S., Perez-Marin, D.: Adapting the design and the use methodology of a pedagogical conversational agent of secondary education to childhood education. In: 2016 International Symposium on Computers in Education (SIIE), pp. 1–6. IEEE, Piscataway, NJ (2016)
7. Pereira, J.: Leveraging chatbots to improve self-guided learning through conversational quizzes. In: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 911–918 (2016)
8. Latham, A.M., Crockett, K.A., McLean, D.A., Edmonds, B., O'Shea, K.: Oscar: An intelligent conversational agent tutor to estimate learning styles. In: International Conference on Fuzzy Systems, pp. 1–8 (2010)
9. Cooper, H.M.: Organizing knowledge syntheses. A taxonomy of literature reviews. Knowledge in Society 1, 104–126 (1988)
10. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. Management Information Systems Quarterly 26, xiii–xxiii (2002)
11. Fettke, P.: State-of-the-Art des State-of-the-Art. Eine Untersuchung der Forschungsmethode „Review" innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 48, 257–266 (2006)
12. vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the Association for Information Systems 37, 205–224 (2015)
13. vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: RECONSTRUCTING THE GIANT: ON THE IMPORTANCE OF RIGOUR IN DOCUMENTING THE LITERATURE SEARCH PROCESS. In: ECIS 2009 Proceedings, Paper 161 (2009)
14. Frohberg, D., Göth, C., Schwabe, G.: Mobile Learning projects - a critical analysis of the state of the art. Journal of Computer Assisted Learning 25, 307–331 (2009)
15. Decker, J., Hobert, S., Schumann, M.: MOBILE AND WEARABLE LEARNING IN ENTERPRISES – APPLICATION SCENARIOS AND TECHNOLOGY SELECTION. In: Gómez Chova, L., López Martínez, A., Candel Torres, I. (eds.) ICERI 2015. Conference proceedings : 8th International Conference of Education, Research and Innovation, pp. 5609–5619. IATED Academy, Valencia (2015)
16. Association for Information Sytems: Senior Scholars' Basket of Journals, https://aisnet.org/page/SeniorScholarBasket (Accessed: 12.07.2018)

17. Cao, J., Zhang, D.: An Intelligent Learning Assistant in Multimedia-based Interactive E-Learning. In: AMCIS 2001 Proceedings, pp. 127–133 (2001)
18. Song, D., Oh, E.Y., Rice, M.: Interacting with a conversational agent system for educational purposes in online courses. In: 2017 10th International Conference on Human System Interactions (HSI), pp. 78–82 (2017)
19. Mikic, F.A., Burguillo, J.C., Llamas, M., Rodriguez, D.A., Rodriguez, E.: CHARLIE: An AIML-based chatterbot which works as an interface among INES and humans. In: 2009 EAEEIE Annual Conference, pp. 1–6 (2009)
20. Coronado, M., Iglesias, C.A., Carrera, Á., Mardomingo, A.: A cognitive assistant for learning java featuring social dialogue. International Journal of Human-Computer Studies, 55–67 (2018)
21. Mikic Fonte, F.A., Nistal, M.L., Rial, J.C.B., Rodriguez, M.C.: NLAST: A natural language assistant for students. In: 2016 IEEE Global Engineering Education Conference (EDUCON), pp. 709–713 (2016)
22. López, V., Eisman, E.M., Castro, J.L.: A Tool for Training Primary Health Care Medical Students: The Virtual Simulated Patient. In: 2008 20th IEEE International Conference on Tools with Artificial Intelligence, pp. 194–201 (2008)
23. Hayashi, Y.: Influence of Social Communication Skills on Collaborative Learning with a Pedagogical Agent: Investigation Based on the Autism-spectrum Quotient. In: Proceedings of the 3rd International Conference on Human-Agent Interaction, pp. 135–138 (2015)
24. Tamayo, S., Pérez-Marín, D.: An agent proposal for Reading Understanding: Applied to the resolution of maths problems. In: 2012 International Symposium on Computers in Education (SIIE), pp. 1–4 (2012)
25. Mejbri, N., Essalmi, F., Rus, V.: Educational system based on simulation and intelligent conversation. In: 2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA), pp. 1–6 (2017)
26. Tegos, S., Demetriadis, S., Karakostas, A.: MentorChat: Introducing a Configurable Conversational Agent as a Tool for Adaptive Online Collaboration Support. In: 2011 15th Panhellenic Conference on Informatics, pp. 13–17 (2011)
27. Goh, O.S., Ghani, M.K.A., Kumar, Y.J., Choo, Y.-H., Muda, A.K.: Massive Open Online Course (MOOC) with Learning Objects and Intelligent Agent Technologies. In: 2014 International Conference on IT Convergence and Security (ICITCS), pp. 1–4 (2014)
28. Wik, P., Hjalmarsson, A.: Embodied conversational agents in computer assisted language learning. Speech Communication 51, 1024–1037 (2009)
29. Graesser, A.C., Cai, Z., Morgan, B., Wang, L.: Assessment with computer agents that engage in conversational dialogues and trialogues with learners. Computers in Human Behavior 76, 607–616 (2017)
30. Kumar, R., Rosé, C.P.: Engaging learning groups using social interaction strategies. In: HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 677–680 (2010)
31. Fadhil, A., Villafiorita, A.: An Adaptive Learning with Gamification & Conversational UIs: The Rise of CiboPoliBot. In: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 408–412 (2017)
32. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77 (2008)

Teaching the Digital Transformation of Business Processes: Design of a Simulation Game for Information Systems Education

Alexander Löffler[1], Borys Levkovskyi[1], Loina Prifti[1], Harald Kienegger[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Munich, Germany
{alexander.loeffler,borys.levkovskyi,prifti,harald.kienegger, krcmar}@in.tum.de

Abstract. The ability to manage business processes in the context of the digital transformation is a key competency that should be addressed in Information Systems (IS) education. One possibility for teaching this competency is through simulation games, but the current ones lack a dynamic view on changing business processes induced by the digital transformation. In this paper, we present the design of a simulation game to teach the digital transformation of business processes within IS education. The game simulates the transformation of a bike manufacturing company to a bike-sharing provider, in which students have to manage changes in the production process in teams during different transformation phases. We argue how our game supports central learning objectives for teaching the aforementioned topic and show the benefit of our game design by running a pilot test with students from IS education using the Systems Usability Scale to evaluate the utility of our implementation.

Keywords: Simulation Games; Game-based Learning; Digital Transformation; Business Processes.

1 Introduction

Digital technologies are becoming a main driver for changes in today's companies. The combination of information, computing, communication, and connectivity technologies fundamentally transforms business strategies, products and services, or business processes of an organization [1]. This implementation of digital technologies goes beyond an increase in the efficiency or effectiveness of current processes and leads to changes in the way a company operates, which is defined as "digital transformation" [2]. A core part of this transformation is to manage the business processes and their changes induced by the current technological developments [3, 4]. To prepare students for their future working environment, these changes require adoption by academia, such as including current trends and technological developments of the digital transformation into the curricula. Based on the previous analysis by Prifti et al. [5], business processes and their change management can be seen as a core competency for

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the future workforce to be competitive in this dynamic environment. Therefore, teaching the digital transformation of business processes is a highly relevant topic for Information Systems (IS) education.

However, in current IS education, specific concepts for teaching the digital transformation of business processes are still missing [6]. As previously stated by Jeyaraj [7], changes of business processes should not only be taught to students in a theoretical manner, but also be experienced by the students. Therefore, Jeyaraj [7] proposes to use a simulated environment for teaching the topic in a more experiential way. Despite the importance of providing enough practical insights into the effects of business process changes, a detailed solution to experience these changes is missing [8, 9]. A promising approach to achieve this are simulation games, which model parts of the reality and simulate dynamic situations based on the model [10]. A simulation game would provide the ability to imitate process change scenarios with students in a classroom environment [11]. However, current simulation games that focus on business processes, such as the simulation game called ERPsim [12], are geared towards static business processes instead and are not designed to simulate dynamic business processes [12, 13]. Hence, there is a need for a simulation game to teach the digital transformation of business processes from a practical perspective.

In order to address this, we present the design of a simulation game to teach the aforementioned topics. Building upon previous findings for teaching business processes in the context of the digital transformation, we provide practical scenarios for experiencing changing business processes. We use a bike manufacturing company as an example and provide a storyline in which the company transforms to a bike-sharing provider. In the course of the transformation, the production process changes, and students playing the game will have to take different actions to manage the process. We argue whether such a game supports the predefined learning objectives when teaching the digital transformation of business processes and evaluate the game by using it in a pilot test with a group of students from IS education. In summary, the main objectives of the study are to design the simulation game and to show its utility for teaching the aforementioned topics in IS education.

The rest of the paper is structured as follows: First, we present related work in the area of simulation games, business processes, and teaching in the context of the digital transformation. Afterwards, we present our research approach for the design and development of the simulation game. As the core of the paper, we present both the didactical and technical design of the implementation. Finally, we present the results of an evaluation with students from IS education using the Systems Usability Scale (SUS) to show the utility of the implementation.

2 Related Work

Simulation games have a long history in various disciplines. While the initial focus was on simulations in the military or economics, it changed to the usage of simulation games for education in business over the last decades [14]. In IS research, this “gamification” becomes more and more important [15] and various examples for simulation games to serve specific goals of IS education exist. For instance, Baume [10] presents a simulation game to teach information management based on the example of the tasks of a Chief Information Officer. The game is based on the structure for simulation games by Kern [16] and consists of a preparation, interaction, and evaluation phase. Thereby, it is shown that the game provides a useful method to teach information management to students. Another example is the work of Léger [12], who presents a simulation game called ERPsim, which teaches the execution of integrated business processes based on a real Enterprise Resource Planning (ERP) system from the software manufacturer SAP. Moreover, Grund & Meier [17] argue that there is a lack of simulation games to teach business information visualization in IS education and a respective game is developed in a following research project [18]. In summary, these games provide useful methods to teach IS content to students. However, they do not contain a dynamic aspect in which students have to adapt to a changing business environment.

In general, handling changing business processes is not a novel topic that arose in the context of the digital transformation, but has been a topic of IS research for a long time. Leading corporations have used the management concept of Business Process Change (BPC) to manage their processes and the resulting changes induced by current developments. Overall, BPC involves any type of process change, either revolutionary or evolutionary [19]. There has been previous research on the impact of BPC on the success of BPC projects [20] and their dynamic complexities [21], but this was geared toward researchers and practitioners. However, it was already mentioned that an integration of the findings into a simulation game could be a fruitful avenue for future research [22]. In the context of the digital transformation, BPC becomes an important concept, as the restructuring of business processes is a core task within the transformation [3, 4]. Therefore, it is also a topic of high interest for IS education [5].

When it comes to teaching BPC, Jeyaraj [7] proposes use of a simulated environment, as this allows students to experience the impact of changing business processes. In this environment, students can simulate the elicitation, modeling, and reengineering of business processes using role-playing activities. For the context of the digital transformation, where BPC becomes a core competency for the future workforce [5], a previous work [6] analyzes requirements for teaching BPC, which include making decisions on the design, execution, and redesign of business processes using a modern ERP system and user interface. In addition, another work compares two prototypes for simulation games that focus on the digitization of business processes [23] and therefore provide basic insights on the development of simulation games in this context.

3 Research Approach

For the development of our simulation game, our research applies the design science approach according to Peffers et al. [24], additionally considering the guidelines from Hevner et al. [25]. In general, we follow their methodological steps that involve the following activities [24]:

1. **Problem Identification and Motivation.** As previously stated, there is a need for more practical approaches such as simulation games for teaching the digital

transformation of business processes in IS education. This fact can be strengthened by considering the experience in IS classes with teaching the digital transformation of business processes. Focus group discussions with lecturers from the IS context, such as the ones conducted by Prifti et al. [5], showed that teaching in the context of the digital transformation requires different learning strategies and curricula, which are not yet covered in traditional IS courses. More specifically, further focus groups highlighted that simulation games offer the possibility to teach behavioral competencies such as teamwork, competition, and decision making [6], which is difficult to address with traditional teaching methods. Therefore, simulation games are highly suitable to teach the digital transformation of business processes to students in IS education and to provide them with a more practical perspective.

2. **Definition of Objectives for a Solution.** As our planned solution is a simulation game for education, the central objective is to reach learning outcomes with our game. In general, students can achieve learning outcomes at three different levels, which are the institutional, program, or course level [26]. In our case, we focus on the development of a simulation game for an IS course and therefore develop learning outcomes on course level. For this, Biggs developed the Structure of the Observed Learning Outcome (SOLO) taxonomy, which defines action verbs for learning outcomes based on five levels [26]. However, as the verbs in the taxonomy are limited, we also considered the revised Bloom's taxonomy as recommended by Anderson & Krathwohl [27] for the definition of learning objectives for the simulation game. In our case, this taxonomy is highly suitable, as it provides action verbs for a range of learning activities [26, 27], which is necessary to cover all tasks in the simulation game. Hence, we define the relevant learning objectives for our solution based on the action verbs from the revised Bloom's taxonomy [27].
3. **Design and Development.** Based on the learning objectives, we iteratively developed a prototype for our simulation game. This prototype serves as design research artifact whose desired functionality is to address the previously defined learning objectives, according to the approach from Peffers et al. [24]. Based on this, the technical architecture was derived and the artifact was created in iterative development rounds. As described by Hevner et al. [25], an iterative development of the artifact helps to get immediate feedback on the design in the construction phase and supports the creation of an artifact that satisfies the requirements that it is meant to solve. In our case, we built a project team that consisted of four people and met every two weeks. From the fourth week on, we had a running prototype of our simulation game that we discussed and extended iteratively. After four months, the final prototype of the simulation game was ready for the first user tests. At this stage, we were able to compare the artifact's functionality with the previously defined learning objectives it should address.
4. **Demonstration.** To demonstrate our artifact, we conducted a pilot user test with 13 students from IS education. During this test, the participants played the game in a 90-minutes session. Therein, we introduced the company and its digital transformation to the players, explained the process they have to manage, as well as the game rules and its user interface. Afterwards, the participants played the game, in which course they worked with the artifact in order to be able to evaluate it later.

5. **Evaluation.** In the previous design and development phase, we argue whether the game has potential to teach the digital transformation of business processes effectively. In order to evaluate this with a quantifiable measure, we used the questionnaire of the SUS by Brooke [28] for the technical aspects of the game. The SUS is an easy to use scale to measure the usability of a system or product, which covers the effectiveness, efficiency, and satisfaction of a user when working with a system. It consists of a ten-item questionnaire, can be used on small sample sizes and still provides reliable results [28]. In our case, with this measure, the students can give feedback on the usability of the simulation game and we can analyze the utility of our artifact for teaching the aforementioned topic.
6. **Communication.** The communication of the results is fulfilled within this article.

4 Game Design

4.1 Learning Objectives

The central goal of our simulation game is to prepare IS students for their future working environment by gaining a deeper understanding of the digital transformation and its effects on a company's business processes. To derive specific learning objectives for this overarching goal, we analyzed learning objectives in the literature for simulation games in IS education and related disciplines. The literature review has been structured according to the methodology proposed by Vom Brocke et al. [29]. We considered databases including IS and education outlets, such as the IEEE Xplore Digital Library, SpringerLink, the ACM Digital Library, and the AIS Electronic Library. In these databases, we searched for learning objectives in IS education that can be applied on simulation games for business processes. Furthermore, we considered previous analyses on competencies in the context of the digital transformation, such as change management, understanding and coordinating workflows, decision making, and teamwork [5]. Based on the results, we built ten learning objectives that define what the students should learn after playing the game. These objectives are defined based on the revised Bloom's taxonomy [27] and are listed below, clustered by the different activities in the simulation game. After playing the game, the students will be able to

1. … explain the described process and its sequence of process steps [30]
2. … illustrate the interdependencies between different process steps [31, 32]
3. … explain the impact of the digital transformation on the business processes [5, 6]
4. … explain the digital transformation in the production process [5, 6]
5. … explain the challenges of BPC in the context of the digital transformation [6, 33]
6. … develop a strategy for profit maximization based on the current market demands [34]
7. … adapt and change the described process according to the developed strategy [35]
8. … analyze and solve problems in a team [5, 6, 36]
9. … make decisions in a changing and competitive environment [6, 30]
10. … critically evaluate one's own decisions and their impact [6, 36]

4.2 Game Scenario

The game scenario is based on previous research in BPC, simulation games, and the digital transformation. Based on the literature, we derived fundamental principles and technical requirements for the implementation of the simulation game. Furthermore, following the design science approach as previously described, we iteratively discussed the game design and its technical requirements. Overall, this led to the following fundamental principles for our simulation game:

- The students work together in teams and compete against other teams in a common market [6]. Collaboration within the team is thus required to manage the game successfully [37, 38].
- The central tasks in the game comprise decisions regarding changing business processes induced by the digital transformation [39]. The students should divide tasks among their teams to analyze the different decisions, discuss them, and come up with a common solution afterwards [40].
- For all decisions, the students work with the interface of a modern ERP system, in this case SAP S/4HANA, in order to analyze the different alternatives and make decisions based on a real ERP system [6, 8, 33, 40].

We embedded the game into the storyline of a bike manufacturing company, as this is a frequently used example for a model company in education to teach business processes [41]. Especially in the context of the digital transformation, this storyline is highly suitable for students in IS education, as it provides a concrete example for an industry affected by the digital transformation, which can still be easily understood without extensive technical knowledge, e.g. about production processes. Our main idea was inspired by findings in the literature that see the Internet of Things (IoT) as an important concept when teaching the digital transformation [42]. Furthermore, the combination of products and services in a rental model, also called Product Service Systems (PSS), is an important topic for teaching changing business processes [43] and therefore is also considered as part of our game.

In our game, we use the scenario that a bike manufacturing company decided to introduce a new bike with IoT components to the market. In a first step, the new bike received positive feedback from test customers, who really liked the new IoT components. Therefore, the company reorganizes the production process to bring these new bikes to the market. However, despite positive customer feedback, the company is experiencing problems with the sales numbers due to the high product price of the IoT bikes. To address this problem, the executives decide to extend the company's business model and to offer bike-sharing services to private customers on the top of the current business. For this, additional components for the bike, e.g. an onboard application to manage bike rentals is necessary. Overall, the tasks within the game comprise the management of the necessary changes within the production process and ensuring that the new business model is running properly.

Overall, the different changes the bike company undergoes in the digital transformation are split up into four phases of transformation, which are depicted in the game in four rounds:

1. **Phase 1**. There is an initial demand of 10.000 bikes on the market. The teams have the possibility to buy additional assembly lines for production of the new IoT bikes and determine a price for their sales.
2. **Phase 2.** The demand increases to 60.000 bikes and the teams can now introduce quality management to improve the production process. Furthermore, new automatic assembly lines can be bought to increase the production capacity.
3. **Phase 3.** The bike company decides to offer bike-sharing services with their IoT bikes. Therefore, the demand increases to 350.000 bikes. The teams have to integrate application development into their process to be able to offer the bike-sharing services. Additionally, they can buy new assembly lines using predictive maintenance, which have a higher capacity and lower rejection rate.
4. **Phase 4.** In the final round, the demand is 400.000 bikes. The teams can introduce additional sales and distribution channels for their offerings and need to establish their offerings in competition with the other teams.

As basic process for the described scenario, we use a bike production process that consists of ten steps. As a foundation, we used a sample production process for bikes as described by Magal & Word [41]. However, we additionally considered the attachment of IoT components to the bike during production, such as sensors, a motor, or a battery [44]. This leads to the business process in Business Process Model and Notation (BPMN) shown in Figure 1.

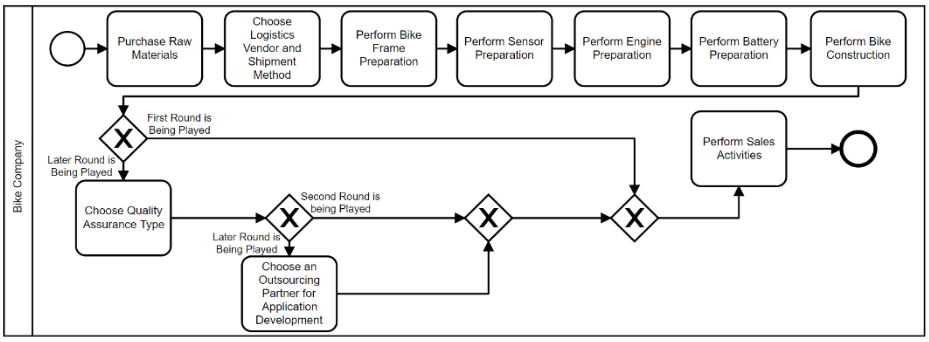


Figure 1. Process Scenario for the Simulation Game

First, there is the purchasing of raw materials, in which the teams can choose between different vendors offering the required materials and can decide which quality they should have. This is followed by logistics, which comprises company-internal logistics and allows the teams to choose between different shipment methods. In the next four steps, which are frame, sensor, engine, and battery preparation, the participants have to choose between different assembly lines. In each transformation phase, new assembly lines can be bought, depending on the current state of the digital transformation. This provides the ability to choose production lines with different capacities and align the production capacity to the overall demand on the market. As the seventh step, the bike construction follows, in which the teams can again select from different types of assembly lines. However, this step is very important, as it determines the final number of bikes the company will produce. Afterwards, there is quality assurance, which does not exist from the beginning, but can be implemented in the second transformation phase and improves the quality of the bikes. The ninth step is application development, which becomes important when the bike company decides to offer bike-sharing services in the third transformation phase. There, the teams have to choose an outsourcing provider for application development in order to manage bike-sharing services with their IoT bikes. Finally, there are the sales activities, which determine the price of the bikes and allow additional decisions that have an effect on the revenue, such as introduction of new distribution channels or sales branches. This last step also determines how many bikes are sold at the chosen price and therefore results in the overall profit of the company.

4.3 Technical Implementation

From a technical perspective, the application is based on the SAP ABAP (Advanced Business Application Programming) technology stack. It consists of three main components, which are the frontend, developed with the JavaScript framework SAPUI5, the backend, developed in ABAP, and an underlying SAP HANA database. The frontend consists of two subcomponents: administration and game. The first one can only be accessed by the instructor and is used to create a new game and to start the round simulation after the player entries have been confirmed. The players interact with the second subcomponent after they have been assigned to an active game. The frontend communicates with the backend for loading of master data, such as material data or vendor data, as well as for game data and round data submission.

In the design of the user interface, we followed the customer-centered approach as described by van Duyne et al. [45] to assure that all the game features are built in an understandable way. First, we have designed our user interface responsively, so that the game can be played using computers, tablets, and mobile phones. We divided the home page into a main section and navigation section, following the navigation bar pattern by van Duyne et al. [45]. From the main area, the players can access every process step shown in the process scenario in Figure 1. A side bar is used for navigating to the round results and to the help page, which describes the game rules and the differences between the rounds. The upper bar contains a button to end the round. Pressing it will disable the ability to change the decisions and will notify the instructor that the team the player is assigned to is finished with the current round.

For every process step, we developed a unified design, consisting of a header area and a decision-making area. Figure 2 shows this on the example of the “Frame preparation” process step. The header area consists of three sections: previous round data, showing the player’s decisions from the last round, current round data, showing what values have been saved for the current round, and cost accounting data, showing the influence of the player’s decisions on the budget, running cost, and total quality of the product. The header area also has the navigation button that returns the player to the home screen. The decision-making area is located under the header and has several input fields that have to be filled in by the player. Each field has a help button that briefly explains its meaning. All changes the players make are also reflected in the live cost accounting section. These values should aid the player through decision making by showing the impact of their decisions without having to save every change. The data needs to be saved only when the player is satisfied with his or her decisions and it is done by pressing the "Confirm" button in the footer of the page. The interface of the game also includes videos that explain the introduced innovations, e.g. robotic quality assurance or predictive maintenance. Located to the right of the live cost accounting, these videos are aimed at providing the player with better understanding of digitalization and automation and thus increasing the didactic value of the game.

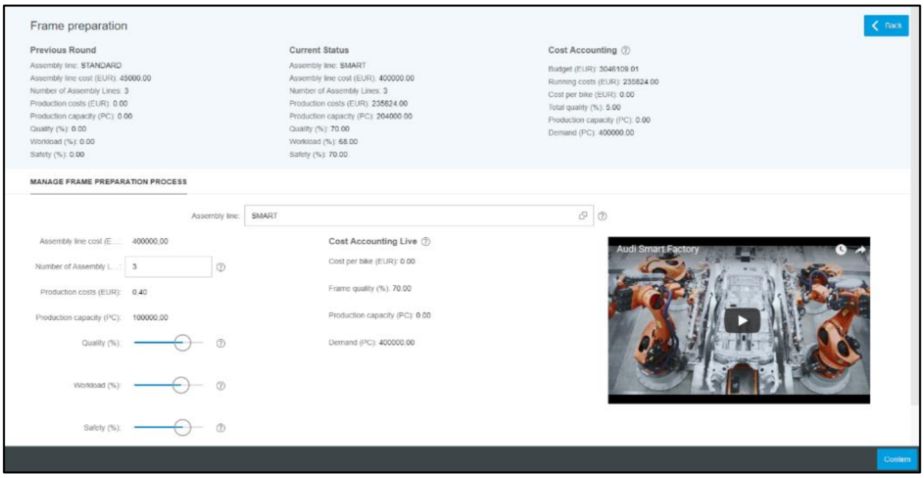


Figure 2. Sample Process Step of the Simulation Game

4.4 Fit Between Learning Objectives and Game Scenario

During the design and development of the game, the main goal was to implement a solution that addresses the learning objectives for teaching the digital transformation of business processes as described in Section 4.1. Table 1 illustrates the fit between the Learning Objectives (LO) and the game scenario from a conceptual perspective.

As illustrated, all learning objectives are covered by different aspects of the game scenario. LO1 and LO2, which focus on explaining and illustrating the business process and its interdependencies, are mainly addressed by the visualization of the basic process scenario in the starting screen of the game. Furthermore, common KPIs for all process steps, e.g. production capacity, create dependencies between the steps that the students have to understand to be successful in the game. LO3 to LO5, which focus on explaining the digital transformation in different contexts, are mainly met by the different activities that are available dependent on the transformation phase. In each phase, new technologies such as smart production lines become available, which have an impact on the efficiency and overall profit in the game. Additionally, new activities have to be handled by the students that are necessary to sustain the profitability, e.g. managing the quality assurance from the second phase on. As general rule, in each round, a new process step appears, which defines the current transformation phase and has to be managed to get a competitive advantage in the game. LO6 to LO8 focus on developing a strategy, adapting and changing the process accordingly, and analyzing and solving problems in a team. These learning objectives are addressed by a changing market demand based on the current round, which is accompanied by further changes regarding the business processes that have to be managed by the students. For this, teamwork is necessary, as new process steps become available in every round and tasks have to be split to react to all changes accordingly. Finally, LO9 and LO10 focus on making decisions in a changing environment and critically evaluating the impact of the decisions. For this, the demand changes in a flexible way based on the number of teams and the participants have to make decisions under time pressure to be competitive. Furthermore, the user interface provides a results screen, which allows students to evaluate their decisions and to reconsider their strategy for the following rounds.

Table 1. Fit Between Learning Objectives and Game Scenario

Learning Objective	*Aspects of the Game Scenario*
LO1	• Process presentation in the preparation phase of the game • Basic process scenario illustrated in the starting screen
LO2	• Interdependencies between the process steps through common KPIs
LO3	• Presentation of IoT components in the preparation phase • Impact of transformation phase on available process steps
LO4	• Impact of technological developments on the production capacity in the game
LO5	• Changes within process steps allow for more process efficiency • New process steps allow for more profit in the game
LO6	• Changing market demands based on the current round • Gameplay allows clear strategy for profit maximization
LO7	• Changes within process steps necessary in every new round • Additional process steps available in every new round
LO8	• Teamwork as fundamental principle of the game • Each team member is required to make decisions
LO9	• Demand changes based on number of teams • Limited time for decision making during the game rounds
LO10	• Results screen for each team to analyze decisions • Different strategies possible to win the game

Overall, regarding the revised Bloom's taxonomy for the definition of learning objectives, all categories are covered with the design of our simulation game. The taxonomy defines six cognitive processes, which are called Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating [27]. While the majority of the learning objectives can be assigned to the processes of remembering (LO1 and LO2) and understanding (LO3-LO5), also applying is covered by our simulation game (LO7). Moreover, the process of analyzing is covered by LO8 and evaluating by LO9 and LO10. Finally, the process of creating is addressed with LO6. However, regarding Biggs' SOLO taxonomy, not all levels are covered by the game.

This taxonomy defines five levels of students' understanding. According to Biggs & Tang [26], four of them can be described with learning objectives, which are called the unistructural, multistructural, relational, and extended abstract level. Regarding the learning objectives of our game, the unistructural (LO1, LO3), multistructural (LO2, LO4-LO6, LO8 and LO9), and relational level (LO7, LO10) are covered. However, the extended abstract level is not addressed, meaning that the current game design is limited to the fact that students get a practical experience within the mentioned topic, but may not be able to transfer their experience to a theoretical level with the current version.

5 Results

5.1 Evaluation Scenario

For the evaluation of our simulation game, we conducted a pilot test with 13 students from IS education. The students were selected randomly from different courses in the IS master program and participated voluntarily. The age of the participants was between 19 and 37 and all of them had a basic understanding of business processes and the term "digital transformation" from previous lectures in their study program. We used a questionnaire including the SUS by Brooke [28] to evaluate the technical design of the game. The SUS is a reliable tool to measure the usability of a system or product and is valid to differentiate between usable and unusable systems [28]. It can be used on small sample sizes and still provide reliable results with regard to the usability of a system [28]. In our case, the SUS is beneficial as it provides a valid result for the utility of our simulation game despite the small number of participants in our pilot test. Based on the results, we can draw conclusions on the technical maturity of the implemented prototype and get a valid measure for the proper design of our game.

Overall, the pilot test took 90 minutes and consisted of three phases, inspired by the structure for simulation games proposed by Kern [16]:

1. **Preparation Phase.** In this phase, first the storyline of the bike manufacturing simulation was explained. The basic process as presented in the previous chapter was explained, as well as the new IoT components that can be added to the bike. Then, the game rules with four rounds, competing teams in a common market, and the fact that the team with the highest profits wins the game were described. Finally, the user interface was introduced to the participants, in order to prepare them for the first round. Overall, this introduction phase took about 20 minutes.
2. **Interaction Phase:** In this phase, the students played the four rounds with the different transformation phases as described in Section 4.2. Before every round, a short introduction to the available decisions was given. Afterwards, the students had ten minutes to make their decisions. After every round, a brief discussion of the intermediate results followed. Overall, the interaction phase took 60 minutes, with about 15 minutes per round.
3. **Evaluation Phase:** After the students played the game and discussed the results, they were asked to fill out an online questionnaire including the ten questions of the SUS. Afterwards, the pilot test was finished.

5.2 Evaluation Results

In general, the students rated the game very positively. The mean score of the SUS was 76.35, with a standard deviation of 11.59. Figure 3 shows the percentile ranking for all SUS scores of the 13 participants.

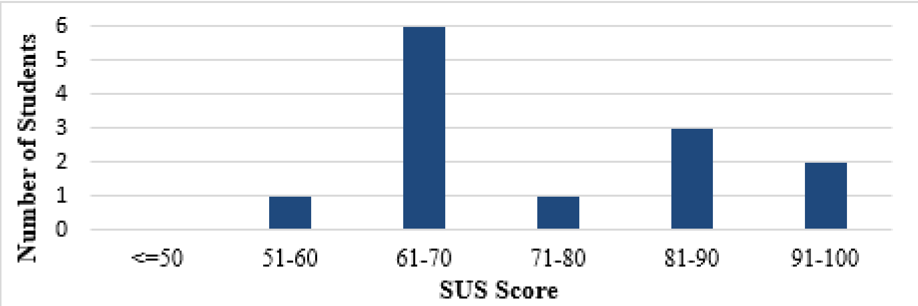


Figure 3. SUS Scores for the Simulation Game

The mean SUS score lies above the average of 68 and therefore is considered a good score [46]. According to the adjective rating score by Bangor et al. [47], the simulation game lies between "good" and "excellent". This indicates that our implementation in this state already has a good technical maturity. Most of the students agree that the simulation game is satisfactory and provides a good usability without needing much help of an instructor, which strengthens the purpose of using the simulation game as a learning system. Moreover, as the SUS not only serves to evaluate the usability, but also the learnability of systems or products [48], we can further derive that our implementation in general facilitates learning. Regarding the previously derived learning objectives, the results from the SUS does not allow drawing conclusions whether the students reached the single learning objectives. However, considering the results as feedback on the technical implementation, we can conclude that the game in general addresses the learning objectives with its technical functionalities. Therefore, the students' feedback through the SUS shows that the simulation game supports the purpose of teaching digital transformation of business processes.

6 Conclusion

In this paper, we have presented the design of a simulation game to teach the digital transformation of business processes within IS education. As basis, we derived learning objectives for teaching the mentioned topic from the literature and defined them for the purpose of our simulation game. The game was built based on the scenario of a bike manufacturing company that undergoes a digital transformation to a bike-sharing provider. We described the game scenario and the technical implementation in detail and argued whether the game design supports the predefined learning objectives. Thereby, we showed that our game is designed to address all level of the revised Bloom's taxonomy and supports the learning experience for students regarding the digital transformation of business processes. However, considering Biggs' SOLO taxonomy, our game design currently lacks the transfer of practical knowledge within the mentioned topic to a theoretical level. Hence, the current game design mainly aims to support students in getting practical experience with the digital transformation of business processes. In order to take the knowledge to a theoretical level, the game design has to be extended including more reflection by the students or a link to related research or literature focusing on further aspects of the digital transformation. Nevertheless, an evaluation using the SUS with 13 students from IS education showed the utility of our implementation from a technical perspective and the support of our technical solution to reach the defined learning objectives. However, a conclusion whether the single learning objectives have been reached by the students is not yet possible. Hence, a more extensive evaluation with a higher number of students can provide further insights on the achievement of the learning objectives. For example, this can include an assessment in which the students have to show their acquired knowledge after they played the game. Moreover, as further evaluation, observations by experienced lecturers may be suitable to measure the learning experience.

In conclusion, with this work, we propose a method for teaching the digital transformation of business processes from a practical perspective. The game can be used by lecturers in their practical courses and thereby extend the learning experience of their students. Furthermore, it provides a recommendation how teaching in the context of the digital transformation can be enriched with more interactive methods. Thereby, we contribute both to the theory of teaching business processes in IS education and to practice by providing concrete guidelines to teach the digital transformation of business processes using a simulation game.

References

1. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.: Digital Business Strategy. Toward a Next Generation of Insights. MISQ 37, 471–482 (2013)
2. Oswald, G., Krcmar, H.: Digitale Transformation. Springer, Wiesbaden (2018)
3. Abrell, T., Pihlajamaa, M., Kanto, L., Vom Brocke, J., Uebernickel, F.: The role of users and customers in digital innovation. Insights from B2B manufacturing firms. Information & Management 53, 324–335 (2016)
4. Nwankpa, J.K., Roumani, Y.: IT Capability and Digital Transformation: A Firm Performance Perspective. ICIS (2016)
5. Prifti, L., Knigge, M., Kienegger, H., Krcmar, H.: A Competency Model for "Industrie 4.0" Employees. 13th International Conference on Wirtschaftsinformatik (2017)
6. Löffler, A., Prifti, L., Knigge, M., Kienegger, H., Krcmar, H.: Teaching Business Process Change in the Context of the Digital Transformation: A Review on Requirements for a Simulation Game. MKWI (2018)
7. Jeyaraj, A.: Business Process Elicitation, Modeling, and Reengineering: Teaching and Learning with Simulated Environments. JISE 21, 253–264 (2010)
8. Monk, E.F., Lycett, M.: Measuring business process learning with enterprise resource planning systems to improve the value of education. Educ Inf Technol 21, 747–768 (2016)
9. Vuksic, V.B., Bach, M.P., Hernaus, T.: Educating Students in Business Process Management with Simulation Games. Int Journal of Bus an Eco Eng 8, 1245–1250 (2014)

10. Baume, M.: Computerunterstützte Planspiele für das Informationsmanagement. Realitätsnahe und praxisorientierte Ausbildung in der universitären Lehre am Beispiel der "CIO-Simulation". München (2009)
11. Seethamraju, R.: Enhancing Student Learning of Enterprise Integration and Business Process Orientation through an ERP Business Simulation Game. JISE 22, 19–29 (2011)
12. Léger, P.-M.: Using a Simulation Game Approach to Teach Enterprise Resource Planning Concepts. JISE 17, 441–448 (2006)
13. Legner, C., Estier, T., Avdiji, H., Boillat, T.: Desiging Capstone Courses in Management Education: Knowledge Activation and Integration Using an ERP-based Simulation Game. ICIS (2013)
14. Geuting, M.: Planspiel und soziale Simulation im Bildungsbereich. Lang, Frankfurt (1992)
15. Blohm, I., Leimeister, J.M.: Gamification. Bus Inf Syst Eng 5, 275–278 (2013)
16. Kern, M.: Planspiele im Internet. Deutscher Universitätsverlag, Wiesbaden (2003)
17. Grund, C.K., Meier, M.C.: Towards Game-based Management Decision Support: Using Serious Games to Improve the Decision Process. MKWI (2016)
18. Grund, C.K., Schelkle, M., Hurm, M.: Architecture and Evaluation Design of a Prototypical Serious Game for Business Information Visualization. 13th International Conference on Wirtschaftsinformatik (2017)
19. Grover, V., Markus, L.M.: Business Process Transformation. M.E. Sharpe, Armonk (2008)
20. Jurisch, M.C., Cuno, J., Palka, W., Wolf, P., Krcmar, H.: An Integrative Model of IT-Enabled Business Process Change: Causal Structures in Theory, Research and Practice. HICSS (2012)
21. Rosenberg, Z., Riasanow, T., Krcmar, H.: A System Dynamics Model for Business Process Change Projects. ICSDS (2015)
22. Rosenberg, Z.: Simulation Approach for Managing and Analyzing Dynamic Complexities in Business Process Change Projects. München (2016)
23. Löffler, A., Prifti, L., Levkovskyi, B., Utesch, M., Krcmar, H.: Simulation games for the digital transformation of business processes. Development and application of two prototypes from the automotive and online retail sector. EDUCON (2018)
24. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. JMIS 24, 45–77 (2007)
25. Hevner, A.R., Salvatore, T.M., Park, J., Ram, S.: Design science in information systems research. MISQ 28, 75–105 (2004)
26. Biggs, J.B., Tang, C.S.-k.: Teaching for quality learning at university. What the student does. McGraw-Hill/Society for Research into Higher Education/Open University Press, Maidenhead (2011)
27. Anderson, L., Krathwohl, D.R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York (2001)
28. Brooke, J.: SUS: A quick and dirty usability scale. Usability evaluation in industry 189, 4–7 (1996)
29. Vom Brocke, J., Simon, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. ECIS 9, 2206–2217 (2009)
30. Chwif, L., Barretto, M.R.P.: Perspectives on simulation in education and training: simulation models as an aid for the teaching and learning process in operations management. Winter Simulation Conference (2003)
31. Forssén-Nyberg, M., Hakamäki, J.: Development of the production using participative simulation games. Two case studies. Int Journal of Prod Eco 56-57, 169–178 (1998)

32. Pittarese, T.: Teaching fundamental business concepts to computer science and information technology students through enterprise resource planning and a simulation game. Journal of Computing Sciences in Colleges 25, 131–137 (2009)
33. Bandara, W., Chand, D.R., Chircu, A.M., Hintringer, S., Karagiannis, D., Recker, J.C., Welke, R.J.: Business process management education in academia: Status, challenges, and recommendations. Communications of the AIS 27, 743–776 (2010)
34. Strecker, S., Rosenthal, K.: Process Modelling as Serious Game. Design of a Role-Playing Game for a Corporate Training. CBI (2016)
35. Fuchsberger, A.: Improving Decision Making Skills through Business Simulation Gaming and Expert Systems. HICSS (2016)
36. Clarke, T., Clarke, E.: Learning outcomes from business simulation exercises. Education + Training 51, 448–459 (2009)
37. Koutsopoulos, G., Bider, I.: Teaching and Learning State-Oriented Business Process Modeling. Experience Report. Enterprise, Bus Proc and IS Modeling, 171–185 (2017)
38. Murawski, M., Bick, M.: Digital competences of the workforce – a research topic? BPM Journal 23, 721–734 (2017)
39. Carillo, K.D.A.: Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. BPM Journal 23, 598–622 (2017)
40. Sutliff, K.: Integrating academics and industry. ACM J. Comput. Doc. 24, 33–38 (2000)
41. Magal, S., Word, J.: Integrated business processes with ERP systems. Wiley, NJ (2012)
42. Gronau, N., Ullrich, A., Teichmann, M.: Development of the Industrial IoT Competences in the Areas of Organization, Process, and Interaction Based on the Learning Factory Concept. Procedia Manufacturing 9, 254–261 (2017)
43. Manlio Del Giudice, P., Zancul, E.d.S., Takey, S.M., Barquet, A.P.B., Kuwabara, L.H., Cauchick Miguel, P.A., Rozenfeld, H.: Business process support for IoT based product-service systems (PSS). BPM Journal 22, 305–323 (2016)
44. Holdowsky, J., Mahto, M., Raynor, M.E. and Cotteleer, M.: Inside the Internet of Things (IoT): A primer on the technologies building the IoT, https://www2.deloitte.com/content/dam/insights/us/articles/iot-primer-iot-technologies-applications/DUP_1102_InsideTheInternetOfThings.pdf
45. van Duyne, D.K., Landay, J.A., Hong, J.I.: The design of sites. Patterns for creating winning web sites. Prentice Hall, Upper Saddle River, NJ (2007)
46. Sauro, J.: Measuring Usability with the System Usability Scale (SUS), https://measuringu.com/sus/
47. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 4, 114–123 (2009)
48. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. International conference on human centered design, 94–103 (2009)

Conceptualizing Immersion for Individual Learning in Virtual Reality

Henrik Kampling[1], Anna Schwarze[1], Oliver Heger[1], and Bjoern Niehaves[1]

[1] University of Siegen, Chair of Information Systems, Siegen, Germany
{henrik.kampling,anna.schwarze,oliver.heger,
bjoern.niehaves}@uni-siegen.de

Abstract. Immersive virtual reality technology (VR) receives more and more attention, especially since the release of the Oculus Rift (development kit 2) in 2016. This technology is not only used in the gaming industry but also in serious contexts such as product design or education. The creation of high immersion is commonly said to be the special characteristic of VR. We consider two perspectives on immersion: firstly, immersion in the task and, secondly, immersion in the technology. Our work focuses on immersion as part of the learning related theory of cognitive absorption to examine the theoretical difference between task and technology immersion in the case of individual learning with immersive VR technology. We conducted an explorative Grounded Theory approach with 10 in-depth interviews based on first-hand experience with a self-developed immersive VR application. We propose theoretical and design implications for how VR can potentially enhance individual learning.

Keywords: immersion, virtual reality, individual learning, grounded theory

1 Introduction

Virtual Reality (VR) technologies are gaining more and more popularity, especially since the market entrance of Oculus Rift (Development Kit 2) in 2016 [1]. Nowadays, there are a couple of virtual reality head-mounted displays (HMDs), such as HTC Vive or PlayStation VR. VR technology is not only used for gaming contexts but in a variety of serious cases, such as product design and manufacturing [2, 3], urban spaces [4], (landscape) architecture and environmental planning [5], travel fair [6], rollercoasters [7], healthcare [8], or education [9]. The latter, for instance, shows an overview of educational perspectives and contexts. Liu et al. propose frameworks for design and implementation of learning [9, part I] and case studies of immersive learning (part II).

Such technologies facilitate new opportunities to improve education. Nowadays, many teaching formats and materials are supported by digital technology, such as massive open online courses, collaborative learning through social media, blended learning concepts, or other e-learning tools. Besides increasing the detachment from local and time restrictions, an essential objective of digitally supported learning is the improvement of individual learning outcome. Recent research in Information Systems (IS) investigates in how far innovative learning strategies [10] and environments [11]

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

can improve individuals or organizational learning outcomes, learning performance and the acceptance of e-learning technologies [e.g. 12].

VR environments based on head-mounted displays (HMD), such as Oculus Rift or HTC Vive, offer high potential to enrich learning experience and achieve better learning outcomes. In an immersive VR, the user is completely surrounded by an enclosing virtual space [13], which requires complex interfaces such as an HMD. Being in an immersive VR, users are entirely "beamed" to a virtual space, in which they interact with the environment using their entire body. Thus, the user is more strongly absorbed by immersive VR than in traditional non-immersive VR [14]. The use of HMDs in combination with the use of controllers allow users to interact with objects in a virtual environment in a more active and deeper fashion [15, 16].

Although we assume that HMD-based VR can significantly increase the immersion of its user compared to traditional screens and, in doing so, increase individual learning outcomes, to the best of our knowledge, no (empirical) research exists explaining interdependencies of immersion VR learning environments. Hence, the meaning and specific conditions of immersion in virtual reality learning environments have to be explored. Consequently, our study is guided by following research question:

RQ: How can immersion be conceptualized in the context of VR technology for individual learning?

Given its explorative nature, our study uses a Grounded Theory approach. In doing so, we use the literature on immersion, VR, and individual learning to get first insights into the subject. Afterwards, we propose our research method in which the interviewees of our study got first-hand experience with an immersive VR-HMD due to the novelty of the technology, i.e. a HTC Vive and a self-developed application. We then present and discuss our findings and provide implications for theory and design.

2 Background on Virtual Reality and Immersion

Virtual reality. In IS research, virtual reality systems are in the scope of interest. Since Oculus Rift (Development Kit 2) entered the market in 2016, the hype surrounding VR technology, particularly HMDs, grew [1]. VR is an interactive, computer-generated three-dimensional environment in which people become immersed [17]. VR applications within such an environment depend on the degree of immersion [13]. On the one hand, non-immersive VR refers most commonly to applications in desktop or laptop computers. On the other hand, immersive VR relates to users who are wearing complex interface technologies (e.g. head-mounted displays) and are completely surrounded by an enclosing virtual space. In addition, VR mostly refers to a single-user interaction in a virtual environment [18] and is typically limited to a single user session of 30 minutes [19]. An overview of current VR technologies, i.e. hardware and software, is provided in Anthes et al. [15].

Current research on VR is manifold and examines diverse topics and fields. A recent study on VR examines virtually high and low experiential products while shopping online [20]. Results show that consumer learn more about products, such as attitude and knowledge, in the virtually experiential high condition than in low condition.

Another study investigates virtual product experience and consumers' product understanding while it focuses on product presentation format and task complexity [21]. Further characteristics of immersive VR are arising through the use of HMDs and the use of controllers [22]. Both technological characteristics allow users to interact, create, and manipulate objects in a virtual environment [23, 24]. Against the background of VR and its origins in the consumer market with focus on gaming, VR is well-suited for applying gamification techniques for teaching cases within architecture and engineering 3D arts because it can help to engage into the learning process [25].

As mentioned above, an advantage of immersive VR is the increase of precision and permission of the visualization of objects and processes which are otherwise difficult or impossible to show in the real world which in addition allows to promote focused experiences, such as for the purpose of learning [14, 24]. Slater and Sanchez-Vives [14], for instance, describe four particular advantages immersive VR can have for educational purposes: First, VR can change abstract settings to tangible settings. For example, geometrical and mathematical concepts can be more easily understood in VR compared to traditional paper and pencil learning [26]. Second, VR settings allow a user/learner to actively engage than just observe how things work. For instance, surgical training, ideally paired with haptic feedback, can profit from practicing instead of observing handles [27]. Third, VR simulations allow to substitute methods that are desirable by teachers but practically infeasible or impossible in reality. If students have to study different elements such as Niagara Falls on week one and Stonehenge on week two, it will be infeasible to visit both places (due to time or resource restrictions) [28]. Fourth, VR environments are able to break the bounds of reality as part of exploration. Within such a virtual setting, physics can be manipulated such as changing gravity or making light speed and biological cell utilizing visible [29].

Immersion in the literature. A central component of VR generated, through the use of HMDs, is immersion [14]. In accordance with Slater [16], a user becomes immersed in a completely surrounding virtual setting (wearing a HMD) so that they can turn in any direction with head movements and motion parallax. Literature defines immersion, firstly, as task immersion (such as interactions and activities) "a mental state of being completely absorbed or engaged with something" [30] and, secondly, as technology immersion (such as software) "the experience of total engagement where other attentional demands are, in essence, ignored" [31]. Both are similar but have different research backgrounds and are measured differently. Within the context of learning, immersion is essential within the flow theory [32], and hence, it is a central component of Cognitive Absorption (CA) in the context of new information technology [31]. Both deal with an individual's mental state of absorption, a feeling of engagement and immersion, including intense concentration, a sense of being in control, a loss of self-consciousness, and a transformation of time [31, 32]. Hence, Flow and CA agree that the role of the flow experience shapes individual attitudes and behaviors when using information technology.

Cognitive absorption and flow have been the object of investigation in many studies. Flow, for instance, has been studied in relation to e-learning environments for higher education [12]. Here, processes of interactivity, imagery, and spatial presence influence flow (directly associated to intrinsic motivation) while the goal is to analyze the user's response towards a learning environment, such as continuance behavior. Another educational and more practical approach focuses on interaction and flow and its impact on e-learning acceptance by nurses [33]. The goal of the study is to analyze interaction factors (such as learner-system, instructor-learner, and learner-learner) and intrinsic (e.g. flow) as well as extrinsic motivators (e.g. TAM constructs, such as ease of use and usefulness) to explain the nurses' behavior of using an e-learning system for continuance intention.

Current studies of CA investigate it in relation to perceived learning in a mobile training scenario [34] and to individual learning in groups through text and video [35]. The first one found out that CA plays a significant role in affecting the deep involvement of users. The second one analyzed peer influenced learning and individual CA on learning outcomes (i.e. satisfaction, perceived understanding, and performance). With regards to a virtual environment, Goel et al. [11] examine the effect of CA on learning in collaborative tasks. They analyzed the effects of CA on perceived learning, learner satisfaction, and task participation in a virtual world. Collaborative learning and cognitive absorption have also been studied in an organizational context including individual learning [36]. The authors focus their work on group-level behavior that, in turn, can reduce (self-assessed) individual learning. Another work focusing on online learning environments investigates social presence and interest as antecedents of CA which, in turn, affects satisfaction [37]. Burton-Jones and Straub [38] present a contextualized model of system usage and individual short run task performance in which CA and deep structure usage are antecedents of system usage. Their results indicate that inappropriate choices of usage measures reduce explanations of performance.

In summary, much research on VR and immersion already exists. With regard to learning, only a couple of studies have examined the actual individual learning but mainly in group or collaborative tasks [11, 36–38]. Nevertheless, these studies do not consider a differentiate view on the construct of immersion, and in this way, flow and CA. On the one hand, immersion is outlined as task or interaction specific [e.g. 11, 32, 36], and on the other hand, it relates to technology [e.g. 31, 38]. Moreover, due to the characteristics of VR-HMD technology (such as the opportunity of enclosed virtual spaces which offer high potential for interactivity, creation, and manipulation of objects), we assume substantial potential in the context of learning. In turn, the meaning of immersion, particularly the differentiation between task/interaction and technology immersion has not been focused yet. In addition, we assume this missing theoretical differentiation as a gap which could be essential for individual learning within VR.

3 Research Design

Methodology. Due to the novelty of VR-HMDs, we are not surprised that no solid theoretical base exists that is able to explain the central construct and the interdependencies of immersion within the context of virtual learning environments. Taking this into account, the study at hand was conducted to get initial insights to comprehend this phenomenon. As a consequence, a deductive approach cannot be used to investigate this context because there is a lack of a reliable body of existing theory to inform extensive a priori theorizing [39, 40]. Due to the explorative nature, we follow an inductive Grounded Theory approach [41–43] to explore and examine the central component immersion in the context of individual learning with VR systems. Grounded Theory is well suitable for exploring theoretical insights [41] and can be characterized along six dimensions [44]: i) development of theory for describing and analyzing the phenomenon of interest; ii) continuous data comparisons against different viewpoints by constant growing analytical and theoretical aspects; iii) a step-by-step coding of data across multiple steps as emerging theory develops; iv) along upcoming differentiating dimensions the theoretical sampling of data; v) the handling of prejudices that prevent relying on any certain theory as a starting point; vi) "an inextricable link between data collection and analysis that incorporates further sampling as part of ongoing analysis and theorizing" [1].

Data Gathering. Against the background of the novelty of immersive virtual reality technologies, such as HTC Vive, Oculus Rift, or PlayStation VR, we conducted the interviews in a virtual reality lab. Here, all interviewees were able to get first-hand experience by using a virtual reality head-mounted display (i.e. HTC Vive). They used a self-developed VR environment (created with the Unreal Engine 4). First-hand experience guarantees that each interviewee is able to answer questions regarding virtual reality systems. The interviews were guided by questions which aim to reveal insights of VR-based learning technologies and of learning (theories) useful for immersive systems.

One major challenge of the interviews results from the fact that VR systems are (to the best of our knowledge) popular but not yet part of everyday life and, as a consequence, largely unknown (particularly with first-hand experience) among non-experts. Therefore, before the interview started, we provided each interviewee with time to use a virtual reality head-mounted display to get first-hand experience. As mentioned above, each interviewee was able to use a self-developed application (c.f. Figure. 1). Each participant was guided to an enclosed room to ensure a quiet and controlled environment without external distractions. Only people who actually used the VR system were interviewed to guarantee that each interviewee was able to answer questions regarding the VR system. Within the self-developed demo, each interviewee was confronted with a scenario to move freely in a virtual room. Here, the starting point was in a room where a door has to be opened by using a key. The key had to be taken with the help of the controller and led to the door, which then opened. The participant had to walk through the door by real walking movements and to follow instructions in order to hand over a parcel, by using the controller, from one point to another, where a virtual avatar was waiting. The system provides feedback to the user through visualizations, i.e. green for correct conduction and red for making a mistake. In case the user made a mistake, he or she could restart the process until they conduct it correctly. Afterwards, each interviewee was given time to explore the virtual setting, such as interacting, playing, or manipulating objects. The aim of the VR demo is a parcel delivery process in which a participant is able to learn a step-by-step process.

For instance, the demo oriented itself within a perspective of a postman who hands over a parcel.

Figure 1. Self-developed immersive VR application

The interviews had an average duration of 67 minutes, whereby 6 male and 4 female persons were interviewed with an average age of 29 years. Within our interview group, we had one ERP consultant, one management consultant, one innovation consultant, one practice nurse, one research associate, one student of teaching, one student of environmental science, two students of information systems, and one student of business administration.

The study at hand follows an interview guide approach because it is more comprehensive and systematic for data collection than a purely conversational interview. Each interview was open-ended to ensure all interviewees can add concerns that we did not cover in our guideline [45]. Each interviewee was able to ask questions about the immersive VR technology while they were using it. Here, we followed the guideline of Darke et al. [45] who suggest to conduct an interview with at least two interviewers. We recorded the interviews to minimize data loss and to provide all answers and insights given by the interviewees. All interviews were fully transcribed.

Data Analysis. For reviewing our interviews, we used MAXQDA 12. Here, we looked for indicators of immersive VR technology, particularly insights into immersion, and specific conditions of individual learning in VR environments. To analyze the indicators, we draw from well-known methods from grounded theory [41–43, 46, 47], i.e. open coding, axial coding, and selective coding. Hence, we had three phases of analysis. These phases were done iteratively by the first two authors of the paper at hand, including phases of independent coding and code-matching to come to a joint result. Within this initial phase, i.e. looking for emerging aspects (open coding), we specifically took care that within the procedure no themes were excluded due to previous experience or prejudice. We then applied axial coding by building clusters of similar codes. The axial coding allowed us to identify different insights relevant for our subject. In a third step, we searched for relations between the insights to better understand the themes and to derive theory implications, we reflected them by matching them with the literature (selective coding). After we conducted 10 interviews, we finished collecting data because no new insights were found (theoretical saturation).

For the following presentation of our findings, we chose quotations which were most suitable to represent the overall findings.

4 Findings

Feeling naturalness within immersive VR. An immersive VR leads to the feeling that you can act with the virtual environment similar to the real world. Acting in VR feels natural for the users, so that a feeling of reality arises. Naturally executed movements by users and that everything behaves in such a way a user expects support this feeling. Their expectations are based on the experiences of the real world. This real feeling also creates barriers that do not exist in the virtual setting but which are perceived by the participant as real, based on their real world experiences.

"The application was very exciting. Firstly, because you really had to interact with the objects as in reality, and you had to be able to rely on the virtual environment reacting exactly as you would have expected. So that the physical effects of the virtual scenario match those of the real world. [...] Ok, I have to say that I really don't have the need to step into the table because I think it will hurt my knee, it is like an invisible barrier. It feels like walking through thick water. Of course, you don't feel anything but it feels like as you can rest on it." (Interviewee 1, ERP consultant)

Another aspect mentioned to enhance a feeling of naturalness relies on acting with real hands rather than controllers. These, in turn, can enhance the engagement within a virtual space. For instance, it was mentioned that connected gloves or similar technologies which visualize hands and fingers within the virtual space could change the kind of interaction with objects and strengthen the feeling of real behavior.

"Only one grabbing mechanism is not enough, there have to be two times five fingers so that you act as if you were using the real hand. For example, in medical cases, you need sensitivity." (Interviewee 10, student of business administration)

"If you pull on a glove, one with some sensors on it would probably be even more natural. […] A hand would be cool. Because now I have operated with the controller, I have seen the controller in the virtual reality. I wish I had seen my hands like the controller. Simply to be able to estimate the distance correctly even if I want to grasp something. If I don't have hands I don't know if I grab at the right place" (Interviewee 2, innovation consultant)

Against the background of traditional teaching formats and teaching in VR, positive and negative perspectives were mentioned. For instance, the reason why a learner can recall information could rely on the medium it was taught in. As a consequence, a VR space can enhance a learner to be engaged to specific content-related details as they exist in a real world environment.

"[…], the details in a lecture are between the lines or are available in texts. The medium of these lectures is the lecturer or the slides, in which I have two sources of attraction from which I must draw the information myself. In VR, I don't have the opportunity to write down information. It was beyond my familiar environment, so my head is naturally much more active to look at details and learn." (Interviewee 4, student of information systems)

Dimensions of Immersion. Almost all interviewees explained that they forgot the outside environment because they had the feeling of being inside the virtual setting and to be cut off from the real world. So far, virtual reality settings allow users to ignore distractions from the real world and to focus on the VR environment. In conjunction with haptic stimuli, the novelty, and variety of (inter)actions delivered by VR technology, users are (again) able to draw their attention to the intended learning task. The part of immersion in technology is, as defined above, the total engagement to a system while all other impressions are ignored (from here we call it *technology immersion*). Due to the technology immersion, the participant is in the position to feel like a part of the game or system, i.e. to dive into the VR system. Through the sensory immersion and interaction within the virtual environment, a feeling of being involved arises. In turn, this would make it difficult for an individual to "escape" the learning context because the VR provides an enclosed surrounding, engages, and transfer the user into virtual space.

"So, you were just in the game [VR system], you couldn't just look at it from the outside, like in a normal computer game. You were just in the game as part of the game." (Interviewee 3, student of teaching)

"Virtual reality is very engaging. It surrounds you and gives you another room. If you combine different media, such as writing, video, or audio, under a certain goal, then you have a great beneficial application. In VR, you have a much wider field of vision than on a PC display. Even if you put three displays side by side, the VR can surround you with a 360-degree environment freely for yourself. […] I think the non-perception of passed time is about the new environment, I have focused on many little things, this took so much time. The virtual room invites you to deal with everything; this shifts the perception of time." (Interviewee 3, student of teaching)

"I know that nothing can happen to me, so I can interact with the virtual reality very interested." (Interviewee 6, research associate)

The psychological perspective of immersion is a mental state in which you are immersed in a certain task or activity (from here we call it *task immersion*). On the one hand, participants are curious and engaged in the content, and on the other hand, there are some hazards that the user loses himself while interacting within VR.

"Yes, the technology is also interesting but more interesting is the presented content." (Interviewee 8, student of environmental science)

"I am already very curious to see what the scenario will look like. With these glasses [the HMD is meant] it will be a very exciting and engaging experience." (Interviewee 2, innovation consultant)

"[…] but the risks that I see in this context is that people get lost in the virtual reality activities." (Interviewee 9, practice nurse)

An emerged issue within the interviews is the loss of the feeling of passed time. In case an individual does not recognize the spent time while being active in the VR, this could be positive for concentration as well as learning. In contrast, if one is not sensing the spent time during VR activities than this could lead to a loss of perceiving the real world.

"I think it has a positive context because the loss of a sense of time is in general a sign for concentration. Therefore, to cut out the surroundings, I would say this could be useful [for learning cases]." (Interviewee 5, student of information systems)

"I think it is a danger that you completely lose yourself in virtual reality activities and you do not notice at all how fast the time passes in reality. You can lose the relation to reality." (Interviewee 9, practice nurse)

Finally, our findings reveal an essential insight about the relationship of task and technology immersion. In the context of problem-solving, a VR system can enhance an individual to cut off real world surroundings and to focus on a specific task. Hence, a higher degree of technology immersion allows a user to be engaged with a mental activity (i.e. a higher task immersion).

"There are situations where you get into a flow. In these situations, you just do and do not think. The VR can probably support that more than if I still have the ambient noises or other things that distract me. The VR can already support this. If I just say, I have a problem here and I want to deal with it now, and then, I put on the headphones and hide the world. VR would be notch up one's performance, which then supports me to ignore the real world and I concentrate on one thing [learning task]." (Interviewee 6, research associate)

5 Discussion

Conceptualizing immersion in virtual learning environments. Against the background of our findings, we conceptualize immersion within the context of individual learning in VR environments as manifold and suitable for educational purposes and the kind of absorption (c.f. Figure 2). Based on these, task immersion and technology immersion are dependent on each other in virtual reality learning environments. Task immersion, as described above, is a mental state of being psychologically engaged with something [30, 32]. Here, interviewee 5 mentioned that the interactivities are engaging and supports one's own concentration. Technology immersion is the total engagement with a technology while other attentional demands are ignored [31]. For instance, in our VR setting, interviewee 3 states that the virtual space surrounds and transport oneself in another room. An antecedents of immersion is naturalness. This can be subdivided into content-related details which enhances task immersion because the knowledge transfer from a VR and real similar setting might be easier than from an unsimilar. The natural feeling of interactions with one's own hands and control of movements in a virtual space rise the sense of naturalness which enhances the feeling of immersion with the technology because the activities are unbiased to interactions in the real world. Individual learning can be related to different purposes in which immersive VR technology can be well-suited to measure performance and learning outcomes. On the one hand, within virtual settings a learner can acquire declarative (theoretical) knowledge, i.e. cognitive outcomes, or they can develop (practical) skills. On the other hand, a VR technology can be satisfying for the learner [10].

There is a vast amount of literature on individual learning outcomes and individual performance measures [e.g. 10, 11, 37, 38] as well as on immersion and its defined categories on task and technology immersion [e.g. 31, 33–36, 48]. As a complement to that, our findings (Interviews 5 and 6) indicate that immersion influences individual learning outcome. Consequently, immersion and its two manifestations are essential and need explanation within the usage of VR technology.

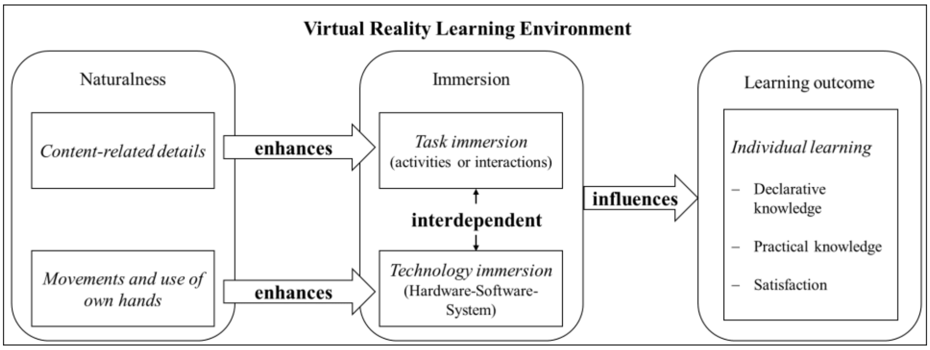


Figure 2. Conceptualization of Individual Learning in Immersive VR

Implications for theory and design. Immersive VR systems have the potential to improve individual learning through the specific characteristics underlying the VR medium. HMDs allow a user to dive into an enclosed virtual space and to interact, to create, and to manipulate objects by using controllers in the virtual environment. Moreover, immersion is essential for individual learning. Our research seeks to conceptualize this construct as a driver for learning outcomes (declarative, practical, and satifsaction). With regard to our RQ ("*How can immersion be conceptualized in the context of VR technology for individual learning?")*, we describe the manifoldness and meaning of immersion in the context of individual learning in VR (c.f. Fig. 2).

Our study provides several theoretical insights. We contribute to previous literature on individual learning outcomes and learning theory by extending the scope of analysis. Individual learning and/or immersion have been primarily investigated within group learning behavior [36], learner satisfaction [37], perceived learning, and task participation [e.g. 11, 34, 35], or from self-regulated learning strategy perspectives [10]. Yet, the use of immersive VR technology within learning contexts has not been focused broadly but already described from a subjective viewpoint [e.g. 9, 14]. We differentiate by interviewing direct user with first-hand experience in which way immersive VR technology can be used, what are preconditions for learning, and by examining the central aspect and the specifications of immersion.

There is an increasing amount of research on immersive VR with an educational purpose [e.g. 26–29]. Within such an environment, real but also abstract and unreal elements can be displayed with a diverse scope of interest. Here, we contribute to the literature by revealing the naturalness, and hence, the feeling of reality by using an immersive VR system that allows a user to develop and gain learning outcomes. Of course, the interviewees perception reveals these different possible outcomes but these are based on their own experience by using an immersive application with an underlying process which a user can learn.

In differentiating the theoretical construct of immersion in case of immersive VR technology for the purpose of individual learning, we contribute to literature by considering immersion in the learning task [e.g. 11, 32, 36, 49] and immersion in the technology [e.g. 31, 35, 37, 38]. Both play, as current literature suggests, an important role for IS educational purposes. From a theoretical perspective, it is worth to explore this phenomenon in more detail because we assume that a higher immersion to technology could decrease the immersion in the task what finally could result in a worse learning outcome. If one is more focusing the technology provided scenario (as a medium that provides the learner content) than the actual learning task, one might argue the learning task is less important or interesting, hence, a learner could be distracted through the technology. In turn, research also has shown that a too high immersion to a task can be detrimental because an intense absorption can lead an individual to ignore contextual cues in which they are immersed [50]. Moreover, if immersion is overemphasized, a learner is more likely to lose their concentration to their performing activity and, therefore, to suffer because of a disconnection from the contextual environment [51]. Here, a learner would rather focus the process than the activity itself. Furthermore, Illies and Reiter-Palmon [52] argue that a potential negative effect on individual learning arises by immersion because the learner might have a closed state of mind, in which they inhibit problem solving and information seeking.

As a consequence, there is a theoretical trade-off for research on immersive VR technology in the context of individual learning. On the one hand, we suggest that researchers should seek to consider this trade-off by focusing both kinds of immersion and taking these considerations into account. On the other hand, immersive VR technology has the potential to be beneficial for learning because such systems can be designed in a way that the antecedents of immersion are increased which, in turn, increases the immersion into the task as well as technology (for instance, design elements of presence [53]). Finally, our findings suggest that there is the opportunity that technology immersion can have a positive effect on task immersion because it can support a learner by ignoring the real world and to concentrate on the task activity.

In accordance with our theoretical implications, the theoretical trade-off can be addressed by designing certain aspects of an immersive VR setting. Users of such systems should be provided with a natural feeling when using the system. For instance, instead of using artificial controller for interacting within a virtual space, Schwind et al. [54] suggest the provision of a "leap motion" technology which could enable a leaner to use their own hands for interactions and perceiving presence. Another example for sensing naturalness refers to the idea that aspects within such a virtual setting has to be considered in a rich way, so that (expected) details are visualized. In contrast, a badly programmed virtual environment will decrease this sense of naturalness.

6 Limitations and Outlook

Our study has several limitations. First, this study is based on interviews, first-hand experiences with 10 participants following a convenient sampling strategy, and was conducted in central Europe. It does not cover how other cultures, certain professions, the elderly, or organizational backgrounds deal with topics around immersive VR technology and individual learning. The study could be extended to other cultures and backgrounds to get a broader perspective and deeper insights into the use of immersive virtual learning environments. Second, although the participants were provided with a VR-HMD experience before the interviews started, all of the participants were previously unfamiliar and inexperienced with such a technology. This first experience could have the effect of overemphasizing a feeling of curiosity or enjoyment with the technology which, in turn, could bias given answers.

As a consequence, this study offers potential for further research. For instance, within an experimental quantitative approach, the differentiation of immersion could be examined to develop a generalizable theory for individual learning (for instance, Jahn et al. [55] suggest in their research-in-progress article a theoretical model). Here, we assume an underlying potential for design theorizing with a focus on certain aspects of immersion to enhance the individual learning outcome. Different immersive VR-based design alternatives could be analyzed with regard to efficiency.

7 Acknowledgements

This article was supported by project ELISE and by a grant provided by the German Federal Ministry of Education and Research (BMBF) (No. 16SV7512).

References

1. Gleasure, R., Feller, J.: A Rift in the Ground: Theorizing the Evolution of Anchor Values in Crowdfunding Communities through the Oculus Rift Case Study. Journal of the Association for Information Systems. 17, (2016)
2. Berg, L.P., Vance, J.M.: Industry use of virtual reality in product design and manufacturing: a survey. Virtual Reality. 21, 1–17 (2017)
3. Choi, S., Jung, K., Noh, S.D.: Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. Concurrent Engineering. 23, 40–63 (2015)
4. Luigi, M., Massimiliano, M., Aniello, P., Gennaro, R., Virginia, P.R.: On the Validity of Immersive Virtual Reality as Tool for Multisensory Evaluation of Urban Spaces. Energy Procedia. 78, 471–476 (2015)
5. Portman, M.E., Natapov, A., Fisher-Gewirtzman, D.: To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Computers, Environment and Urban Systems. 54, 376–384 (2015)
6. Gibson, A., O'Rawe, M.: Virtual Reality as a Travel Promotional Tool: Insights from a Consumer Trael Fair. In: Jung, T. and tom Dieck, M.C. (eds.) Augemented Reality and Virtual Reality. pp. 93–108. Springer Berlin Heidelberg, New York, NY (2017)
7. Jung, T., tom Dieck, M.C., Rauschnabel, P., Ascencao, M., Tuominen, P., Moilanen, T.: Functional, Hedonic or Social? Exploring Antecedents and Consequences of Virtual Reality Rollercoaster Usage. In: Jung, T. and tom Dieck, M.C. (eds.) Augemented Reality and Virtual Reality. pp. 247–258. Springer Berlin Heidelberg, New York, NY (2017)
8. Stone, R.J.: Blending the Best of the Real with the best of the Virtual: Mixed Reality Case Studies in Healthcare and Defence. In: Jung, T. and tom Dieck, M.C. (eds.) Augmented Reality and Virtual Reality. pp. 277–294. Springer Berlin Heidelberg, New York, NY (2017)

9. Liu, D., Dede, C., Huang, R., Richards, J.: Virtual, Augmented, and Mixed Realities in Education. Springer (2017)
10. Wan, Z., Compeau, D., Haggerty, N.: The Effects of Self-Regulated Learning Processes on E-Learning Outcomes in Organizational Settings. Journal of Management Information Systems. 29, 307–340 (2012)
11. Goel, L., Johnson, N.A., Junglas, I., Ives, B.: How Cues of What can be done in a Virtual World Influence Learning: An Affordance Perspective. Information & Management. 50, 197–206 (2013)
12. Rodríguez-Ardura, I., Meseguer-Artola, A.: E-Learning Continuance: The Impact of Interactivity and the Mediating Role of Imagery, Presence and Flow. Information & Management. 53, 504–516 (2016)
13. Mills, S., Noyes, J.: Virutal Reality: An Overview of User-related Design Issues. Interacting with Computers. 375–286 (1999)
14. Slater, M., Sanchez-Vives, M.V.: Enhancing Our Lives with Immersive Virtual Reality. Frontiers in Robotics and AI. 3, (2016)
15. Anthes, C., Garcia-Hernandez, R.J., Wiedemann, M., Kranzlmuller, D.: State of the art of virtual reality technology. Presented at the March (2016)
16. Johnson-Glenberg, M.C.: Embodied Education in Mixed and Mediated Realities. In: Liu, D., Dede, C., Huang, R., and Richards, J. (eds.) Virtual, Augmented, and Mixed Realities in Education. pp. 193–218. Springer (2017)
17. Wexelblat, A. ed: Virtual Reality Applications and Explorations. Academic Press Professional, Inc., San Diego, CA, USA (1993)
18. Schultze, U.: Embodiment and Presence in Virtual Worlds: A Review. J Inf Technol. 25, 434–449 (2010)
19. Nilsson, A., Axelsson, A.-S., Heldal, I., Schroeder, R.: The Long-term Uses of Shared Virtual Environments: An Exploratory Study. In: Schroeder, R. (ed.) The Social Life of Avatars. pp. 112–126. Springer London (2002)
20. Suh, K.-S., Lee, Y.E.: The Effects of Virtual Reality on Consumer Learning: An Empirical Investigation. MIS Quarterly. 29, 673–697 (2005)
21. Jiang, Z. (Jack), Benbasat, I.: The Effects of Presentation Formats and Task Complexity on Online Consumers’ Product Understanding. MIS Quarterly. 31, 475–475 (2007)
22. Seibert, J., Shafer, D.M.: Control mapping in virtual reality: effects on spatial presence and controller naturalness. Virtual Reality. 22, 79–88 (2018)
23. Khalifa, M., Shen, N.: System Design Effects on Social Presence and Telepresence in Virtual Communities. In: Twenty-fifth International Conference on Information Systems (2004)
24. Martín-Gutiérrez, J., Mora, C.E., Añorbe-Díaz, B., González-Marrero, A.: Virtual Technologies Trends in Education. EURASIA Journal of Mathematics, Science and Technology Education. 13, (2017)
25. Villagrasa, S., Fonseca, D., Durán, J.: Teaching Case: Applying Gamification Techniques and Virtual Reality for Learning Building Engineering 3D Arts. In: Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality. pp. 171–177. ACM, New York, NY, USA (2014)
26. Hwang, W.-Y., Hu, S.-S.: Analysis of Peer Learning Behaviors Using Multiple Representations in Virtual Reality and their Impact on Geometry Problem Solving. Computers & Education. 62, 308–319 (2013)
27. Müns, A., Meixensberger, J., Lindner, D.: Evaluation of a Novel Phantom-based Neurosurgical Training System. Surg Neurol Int. 5, (2014)

28. Lin, H., Chen, M., Lu, G., Zhu, Q., Gong, J., You, X., Wen, Y., Xu, B., Hu, M.: Virtual Geographic Environments (VGEs): A New Generation of Geographic Analysis Tool. Earth-Science Reviews. 126, 74–84 (2013)
29. Lartigue, J., Scoville, T., Pham, M.: Promoting K-8 Learning using Oculus Rift: Employing Virtual Reality to Increase Learning Outcomes in Elementary Biology. Presented at the E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education October 27 (2014)
30. Dede, C., Jacobson, J., Richards, J.: Virtual, Augemented, and Mixed Realities in Education. In: Liu, D., Dede, C., Huang, R., and Richards, J. (eds.) Virtual, Augmented, and Mixed Realities in Education. pp. 1–18. Springer (2017)
31. Agarwal, R., Karahanna, E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage. MIS Quarterly. 24, 665–694 (2000)
32. Csikszentmihalyi, M.: Flow: the psychology of optimal experience. Harper & Row, New York (1990)
33. Cheng, Y.-M.: Exploring the Roles of Interaction and Flow in Explaining Nurses' E-Learning Acceptance. Nurse Education Today. 33, 73–80 (2013)
34. Reychav, I., Wu, D.: Are Your Users Actively Involved? A Cognitive Absorption Perspective in Mobile Training. Computers in Human Behavior. 44, 335–346 (2015)
35. Reychav, I., Wu, D.: Mobile Collaborative Learning: The Role of Individual Learning in Groups Through Text and Video Content Delivery in Tablets. Computers in Human Behavior. 50, 520–534 (2015)
36. Magni, M., Paolino, C., Cappetta, R., Proserpio, L.: Diving Too Deep: How Cognitive Absorption and Group Learning Behavior Affect Individual Learning. Academy of Management Learning & Education. 12, 51–69 (2013)
37. Leong, P.: Role of Social Presence and Cognitive Absorption in Online Learning Environments. Distance Education. 32, 5–28 (2011)
38. Burton-Jones, A., Straub, D.W.: Reconceptualizing System Usage: An Approach and Empirical Test. Information Systems Research. 17, 228–246 (2006)
39. Myers, M.: Qualitative Research in Information Systems. Management Information Systems Quarterly. 21, (1997)
40. Venkatesh, V., Brown, S., Bala, H.: Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. Management Information Systems Quarterly. 37, 21–54 (2013)
41. Strauss, A., Corbin, J.M.: Basics of qualitative Research: Grounded Theory Procedures and Techniques. Sage Publications, Thousand Oaks; CA (1990)
42. Urquhart, C., Lehmann, H., Myers, M.D.: Putting the 'Theory' Back Into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. Information Systems Journal. 20, 357–381 (2010)
43. Urquhart, C., Fernandez, W.: Using grounded theory method in information systems: the researcher as blank slate and other myths. Journal of Information Technology. 28, 224–236 (2013)
44. Birks, D.F., Fernandez, W., Levina, N., Nasirin, S.: Grounded Theory Method in Information Systems Research: Its Nature, Diversity and Opportunities. European Journal of Information Systems. 22, 1–8 (2013)
45. Darke, P., Shanks, G., Broadbent, M.: Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism. Information Systems Journal. 8, 273–289 (2002)
46. Glaser, B., Strauss, A.: The Discovery of Grounded Theory. Weidenfeld and Nicholson, London (1967)

47. Urquhart, C., Fernandez, W.: Grounded Theory Mehtod: The Researcher as Blank Slate and Other Myths. Presented at the Proceedings of the 27th International Conference on Information Systems, Milwaukee, Wisconsin (2006)
48. Rodríguez-Ardura, I., Meseguer-Artola, A.: What Leads People to Keep on E-Learning? An Empirical Analysis of Users' Experiences and Their Effects on Continuance Intention. Interactive Learning Environments. 24, 1030–1053 (2016)
49. Csikszentmihalyi, M.: Play and Intrinsic Rewards. In: Flow and the Psychology of Discovery and Invention. pp. 135–153. HarperCollinsPublishers (1996)
50. Ormrod, J.E.: Essentials of Educational Psychology. Upper Saddle River, New Jersey: Prentice Hall (2008)
51. Sinha, I.: The Cybergypsies: A True Tale of Lust, War, and Betrayal on the Electronic Frontier. New York: Viking (1999)
52. Illies, J.J., Reiter-Palmon, R.: The Effects of Type and Level of Personal Involvement on Information Search and Problem Solving1. Journal of Applied Social Psychology. 34, 1709–1729 (2004)
53. Kampling, H.: The Role of Immersive Virtual Reality in Individual Learning. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018)
54. Schwind, V., Knierim, P., Tasci, C., Franczak, P., Haas, N., Henze, N.: "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 1577–1582. , NY, USA (2017)
55. Jahn K., Kampling H., Klein H. C., Kuru Y., Niehaves B. Towards an Explanatory Design Theory for Context-dependent Learning in Immersive Virtual Reality. In: Proceedings of the 22th Pacific Asia Conference on Information Systems (PACIS 2018), Yokohama, Japan (2018)

Designing a Flipped Classroom Course – a Process Model

Kristin Vogelsang[1], Alena Droit[2], Kirsten Liere-Netheler[1]

[1] Osnabrück University, Organization and Information Systems, Osnabrück, Germany
{kristin.vogelsang,kirsten.liere-netheler}@uos.de
[2] Osnabrück University, Management Support and Information Systems, Osnabrück, Germany
alena.droit@uos.de

Abstract. Digital learning has become more than just a trend in the modern world. Blended learning concepts are well established in different areas of application. An important concept in this domain is the so-called flipped classroom. This approach repurposes class time to focus on application and discussion, while the acquisition of basic knowledge will happen at home, enabled by online lectures. In the past, research demonstrated and discussed the advantages of flipped classroom concepts within case studies. Still, standardized guidelines for the development of flipped classrooms are rare. However, it is necessary to learn from the past to improve future education. Thus, we analyzed reviews on flipped classroom research and used these to develop a generic process model for the realization of flipped classroom concepts. The model is based on phases taken from project management, which help to structure the procedure and associated tasks.

Keywords: Flipped Classroom, Course Development, Project Management, Process Model, Checklist

1 Introduction

The concept of a flipped classroom (FC), also known as inverted classroom, has gained rising attention over the last few years. It was primarily described by Bergmann and Sams in 2006 [1]. The number of publications as well as practical implementations are still increasing [2], [3]. A common understanding of the flipped classroom is that the activities of attendance time and time outside the classroom are switched [4]. Bishop and Verleger understand "the flipped classroom as an educational technique that consists of two parts: interactive group learning activities inside the classroom, and direct computer-based individual instruction outside the classroom." [3] The impacts of using this concept are widely discussed. Even though some approaches exist which conclude that FC does not improve class performance, compared to traditional lectures [5], [6], the majority of research results confirm positive impacts on student outcomes (like performance and satisfaction) as well as class participation when self-paced learning is in focus [3], [7-9].

Until now, only small parts of lectures are held as FC. One reason for this is a lack of knowledge about the design of these courses. Structural research providing an

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

overview of the topic is rare [10]. The dominating part of research available is case-based which leads to a “siloed” character of the research field, missing systematic approaches [2], [11]. This paper aims to present a process model for the course development from a lecturer’s perspective. We understand a process model as a guideline including basic tasks and milestones, which are successively being processed and are striving towards a clear goal. Theoretical guidelines can help to design and use FC and are recommended to be used for implementation [3]. To proceed systematically, we align the necessary steps to project management phases. To identify tasks and challenges associated with each phase, we conduct a review of literature reviews about FC that is later enriched by a forward and backward search. Additional information on important to do’s and possible questions are provided.

This paper contributes to research and practice by using a structural approach from another field of knowledge to systematize FC research. It also helps lecturers to design an FC. The focus lies on the development of a useful guideline for practitioners. In the next chapters, we present the research method and summarize findings from FC reviews, which are used for the development of the process model that is presented in chapter 4. We conclude by summarizing and discussing the findings on the phases. Lastly, the limitations of our paper and an outlook for further research are shown.

2 Method

We regard the conceptualization of an FC class as a process that follows all major project management phases (initialization, planning, execution and closing) according to the Project Management Body of Knowledge (PMBok) [12]. Using this concept is unusual as several teaching designs and concepts already exist [13]. Teaching and traditional instructional designs (e.g., ADDIE [14]) include aspects of the competences, group of learners, and teaching subjects [13]. However, their dominating parts focus on pedagogical issues rather than on processual aspects [15]. The process-oriented step by step guideline is (especially in the field of FC-Design) still underrepresented [10]. Therefore, we chose a project management guideline and added pedagogical insights, when appropriate. We assume that this approach is easy to apply and understand due to the few phases involved.

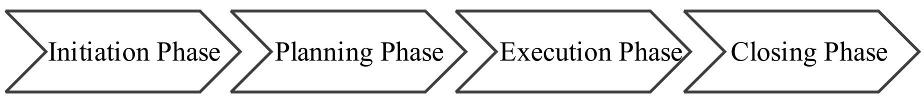


Figure 1. Project Management Phases

In general, different guidelines to define project phases exist and the number of phases varies [12]. Nevertheless, generic theoretical definitions can be applied to different kinds of projects. They all have in common that the phases are sequential and the degree and uncertainty are greatest in from the outset when stakeholders can best be involved [12]. We find many characteristics that are typical for projects within the creation of an FC course as the concept is new, of limited resources and limited time

[12]. Using a theoretical perspective, we define four key phases to develop, implement, proceed and evaluate an FC course (see figure 1). During the initiation phase, the idea of the project comes up and has to be evaluated. Risks and impacts are considered to prepare the decision about the project execution. In the second phase, the planning, required resources are identified and a plan for time, costs and performance is developed. The third phase is often presented in two sections (testing and execution). It describes the integration of products or services designed in the project. The closing phase includes an evaluation of the project and its output.

These phases are used to define activities for developing and implementing a flipped classroom. The activities are identified by conducting a literature search, that reflects the current knowledge base of FC [16]. To get the best possible overview, we decided to conduct a meta-review of existing literature reviews. The findings of the individual reviews can be compared and contrasted [17]. The reviews are used to (1) give an overview of the field of FC research, (2) find activities for the FC development phases and (3) start a forward and backward search. The forward and backward search leads to FC case studies describing the performed activities. The process model is developed in iterative steps. In each step, the identified activities are collected, discussed and assigned to the phases. Figure 1 shows how the research methods and the structure of the paper are linked.

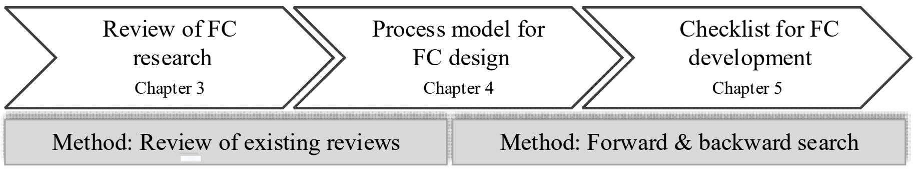


Figure 2. Research process

Our search was conducted in the following databases: Web of Science, Science Direct, Google Scholar, ERIC, AISNET and Scopus. We combined two research strings. The first describes FC to identify articles on the topic: „flipped classroom“ or „inverted classroom“ or “flip teaching”. The second research string is used to limit the results to review articles. Terms used are “review” or “state of the art” or “state-of-the-art” or “meta”. Both research strings are combined with an “and” function. We regard our work as a meta-analysis in the broadest sense [18], as we did not statistically analyze the databases.

The search resulted in 70 hits. Duplets and mere case-descriptions were sorted out which led to a total number of 21 articles, published between 2013 and 2018. We examined the focus of the reviews as well as the major findings. Moreover, we identified the kind of learners, the learning context and how the FC was implemented. The results of this analysis are used as a base to describe the state-of-the-art in FC research (see chapter 3). In a second step, we used the reviews to identify specific topics and tasks of the process of developing the FC. Articles, which showed evidence for at least one of the process' phases, were analyzed in-depth. Consequently, we enriched the reviews by findings from further literature in the field of FC. To decide if the papers are useful for identifying tasks for the phases, we used the following selection criteria:

a) Tasks during FC implementation must be mentioned or at least described; b) Clear alignment to the process phases of the development, when tasks were performed; c) The case was assessed as a representative example (no unusual designs); d) The articles had a clear relationship to FC as a concept described by Bergmann and Sams in 2012 [1]; e) The results described a positive influence of the FC. We are referring to an FC term in a broader sense, as articles with additional traditional lecture content have also been included.

3 Findings in Flipped Classroom Research

FC is a highly contemporary subject with a steady increase in publications. More than half of the identified reviews were published within the last two years, eight of them in 2018. Within our literature research, we identified 21 reviews in total. Six of them are meta-studies, e.g., [9], [7]. Eight reviews focus solely on teaching in health care, e.g., [6], [19] and three reviews examine FC courses in engineering [20–22]. Most reviews have been published in the Anglo-American region and focus on the US teaching system. One exception is a meta-study by Tan examining the effectiveness of FC in China. The author concludes that the satisfaction with FC in the Chinese study is significantly higher than in the Western countries and attributes this mainly to the different (teaching) culture in China, which traditionally entails less interaction with the students and limited exchange of opinions [23]. Regardless of the geographical location, several authors also observe a concentration of FC approaches on STEM (science, technology, engineering, math) and health students [2], [21], [9].

In most of the examined FC courses, videos are used to convey knowledge before the face-to-face session, allowing students to progress according to their own learning pace [24]. The attendance time is mainly used to apply that knowledge and to encourage group work and discussions [21]. Within the scope of digitization through FC, it is possible to introduce Learning Analytics (LA) which is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning" [25]. Within the FCs, it can be used to enhance the development of targeted learning materials, monitor the success of in-class activities [26] and enrich the final evaluation [27].

The majority of reviews state that FC approaches have positive effects on the success of a course compared to traditional lectures. These include increased overall performance, more cooperative learning and increased student satisfaction as the format supports discussions between students and teachers [7], [20], [28], [29]. Furthermore, better learning habits and positive attitudes are observed [21]. Nevertheless, some authors criticize the lack of control groups in many studies and state that the results of some studies are not statistically significant [6], [34]. Due to the different design possibilities of the FC courses, comparability is difficult. There are also a few results which state that student outcomes are not better in FCs compared to traditional methods [6]. Accordingly, new approaches for the evaluation of FC courses which unequivocally prove the success of FC are needed [5], [19]. Moreover, the high initial costs and set-up times incurred when implementing an FC course, especially for the lecturers, are not yet sufficiently investigated [21], [30].

In sum, most reviews show the results of delimited case studies that focus strongly on individual disciplines. This leads to a siloed and perhaps anecdotal knowledge in the research field without any systematic approaches. Therefore, a more general systematic examination of current research is necessary [2], [11].

4 Process Model for the Design of a Flipped Classroom

In the following chapter, we will describe the activities to be carried out to run an FC course. Figure 3 shows the project phases and the respective milestones to mark the (intermediate) results of each phase. The *milestones* will be described in detail below within the following chapters. This summarizing overview provides the structure to understand the separate actions taken in each phase.

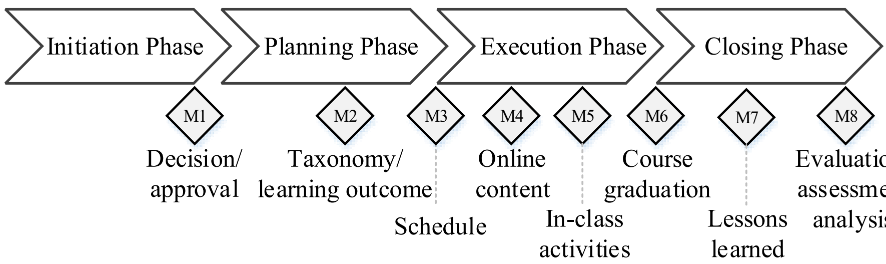


Figure 3. Milestones

4.1 Initiation Phase

The initiation phase aims to prepare a basis for deciding whether a course should be redesigned (or newly created) according to the FC method. Reasons for the redesign or creation could be the discontentment of teachers or students. The difficulties and problems of the current form of the lecture are investigated. Then solution proposals should be created, in this case, the transformation of the teaching form by the implementation of FC. The FC method can, for example, increase the student's motivation, performance, attendance and interaction during face-to-face sessions [20], [31]. It is useful to log the goals of the reorganization for the later evaluation. Once the goals which should be achieved by the FC have been determined, the time, staff and financial expenses of the implementation can roughly be estimated [12]. It should be checked whether necessary resources for a project team are available and to what extent costs for new acquisitions (e.g., learning management systems (LMS), video equipment) will be incurred [21], [30]. The affected stakeholders, as well as their benefits, challenges and barriers [12], must be identified. The stakeholders are the lecturers, students and the organization itself. All groups of stakeholders must be provided with the information needed. Teachers should familiarize themselves with the concept and may have to deal with a lack of skills and resources [30], [32].

Implementing the FC requires much effort, especially at the beginning, as the entire course has to be arranged in the FC format [21], [33]. Moreover, technical obstacles like cutting and uploading videos or the provision of self-learning tests on online platforms must be overcome [30]. Students can be actively involved in the redesign of the course by incorporating their feedback and ideas. However, according to Bishop, not all students are very fond of FC [3]. Students usually need more time to prepare for an FC course than to follow up on traditional lectures. It is therefore an important task to motivate the students to prepare the classes. This requires the implementation of appropriate structures. The gamification approach can be beneficial if it corresponds to the motivational structures and preferences of the students [34]. Short quiz questions to enter the next content or competitions between students can be useful to make self-study of class content easier and more appealing. [30], [38]. Here too, technical and skill related hurdles must be overcome as soon as possible [32]. Apart from teachers and students, stakeholders within the organization may also be addressed, such as administrative staff or IT support [35]. At some universities, there are competence centers for virtual teaching or higher education didactics, which accompany the conversion of courses and provide expert advice. Additionally, most organizations have a learning management system that can be used to make the multimedia files available to learners. During the initiation phase, a decision should also be made as to whether Learning Analytics is to be used and whether the technical and personnel requirements are met. The administration and the affected students must be informed and consent to the use of LA within the framework of existing data protection laws.

At the end of the initiation phase, it has to be pondered whether the benefits of FC merit the time and financial investments. If so, the phase ends with the milestone *decision (M1)* to implement an FC course.

4.2 Planning Phase

When the decision is made to implement the FC, the second phase begins. The goal is to plan the flipped classroom in general (adjustments to the curriculum, set the timetable, etc.) and in detail (design and tuning of the lectures). A lack of time is one of the most threatening challenges [30], [36]. Thoughtful planning is essential [11] for the success of the FC. The conception of the FC requires several resources [36], in addition to the pedagogical organization of the course, the instructor needs support to create virtual content and quiz questions. A FC can be conducted without supporting structures within the school. It is advantageous for the planning and implementation of a FC to build a team of people in order to share tasks and exchange experiences as well as results. For the training of the teachers, there is only limited structured information available about the pre-learning activities and duties [36]. Balan recommends the introduction of learner groups to provoke higher student motivation and to focus on individual learning motivation and outcome reviews [37].

Teachers need to get and give information before the execution about needed adjustments in curricula as the course (format) could also impact these [38]. It should also be checked if possible methods used for the FC meet the university's requirements, e.g., examination regulations. If the contents or the form of exams (e.g., digital exams) are to be changed in the context of the conversion, compliance must be ensured. This is also closely related to the *learning outcomes* which the lecturer wants to achieve within the course. Learning *taxonomies (M2)* are useful to structure the goals of the course [29], [39] and are therefore an important milestone in order to implement an FC. They are applied to split the content into reasonable sections and enable the tuning between the online and in-class courses [39]. There are different designs that represent a full flip or partial flip [3]. Furthermore, they reflect the different learning levels [40]. As the online-videos often cover basic contents, the in-class courses can be used for application, discussion, problem-solving and collaborative learning [41]. In this context, it is important to identify the group of students who will be taught. Teachers can develop FC classes for pupils, undergraduates [41], higher education students [11], and specific professional groups [8] from different disciplines.

Furthermore, different learning types need to be considered [3]. The development of the FC can be improved if the lecturer is aware of the diversity of the class. Gender differences can affect students` perceptions and learning outcomes [42]. To convince students of the new method, the concept of FC should be explained beforehand, including the content, goals and the procedure [30]. For the quality of the learning, the design of in-class and out-of-class activities is very important [22], [32]. Pre-class online lectures are of great value if they provide students with the basic knowledge to proceed with interesting in-class actions [36]. For the time out of class mostly prerecorded video lectures, podcasts or screencasts are used [3], [22], [43]. Herreid et al. surveyed FC teachers, with the result that most of them either chose sources like the Kahn Academy for precasted videos or produced the videos themselves, using programs like Camtasia or apps like Educreation [44]. As students can get easily distracted [45], it is recommended to use videos with a length of ten to 20 minutes [22]. The materials for the videos (e.g., slides) need to be planned beforehand [11] and produced step by step [46] before providing them to the students. For this purpose, subject areas have to be divided into several online contents [47]. The development of online courses is time-consuming. Researchers calculate the expenditure with approx. 1-2 hours per unit [33] and altogether 100 hours per course [48], whereby there is also the possibility of using already existing lectures [48]. The videos can be posted on platforms like YouTube, iTunes U, or on LMS like Blackboard and Moodle [44]. Lecturers should stay in contact with the IT support [30], [35] to guarantee easy access for the students [49]. It has proved to be beneficial to use existing technologies rather than developing new ones [50], [51]. As questions cannot be asked immediately [52] forums can enable discussions on the video content [53]. Regular quizzes that mirror the video content help to reduce distraction and ease the preparation for in-class activities [22], [45]. Out-of-class activities can be complemented by homework, pre-readings, automated tutoring systems or supplemental videos [11], [43]. Besides the planning of online lectures, lecturers need to decide which methods should be used in-class and prepare materials if needed. The pre-class preparation of an FC is much more time-consuming and complex than in traditional courses.

If the lecturer wants to use Learning Analytics to monitor student's activities, it has to be decided which data should be analyzed (e.g., trace data of the LMS like activation of course videos or solved online assessments), how the data is analyzed and how the results used [26]. Apart from FC specific planning, regular activities like scheduling in-class time and the reservation of rooms are necessary. The transition to the third phase takes place when the planned course starts. This is only possible when the milestone of *scheduling (M3)* is finished so that the course has a general structure.

4.3 Execution Phase

The actions proceeded during the term are subsumed within the execution phase. The phase mainly aims at the *supply of the video tutorials (M4)* and the *proceeding of the in-class lectures (M5)*. This phase is the one with the highest interaction between the students and the lecturer. This means the rules of the FC must be communicated beforehand [37], as students are less satisfied with unclear instructions and unknown situations [2]. The FC success depends mainly on the student's compliance [1], [30].

In-class courses can be designed as homework, quizzes, lectures, small group activities, presentations (e.g., case-based, student) and discussions (e.g., team-based, panel or expert-led) [3], [11], [43]. The activities chosen are very important as they differ in their effectiveness and conditions needed [43]. If the major goal is to enrich materials in class, more lecturer-oriented activities such as teacher-led discussions are useful. Interactive group work can be more suitable for the application of the material. Moreover, the course size has to be considered. While videos can be used for different group sizes, including large groups [54], the attendance time needs to be planned more carefully for larger groups, for example by forming smaller groups and/or using peer-learning [55], [56]. Attendance time activities are often accompanied by smartphone apps, pair-and-share activities or clicker assessments for immediate feedback to bridge expectations [37], misunderstandings and knowledge gaps [11]. Furthermore, accompanying in-class assessments make it possible to test previous knowledge and ensure quality [51]. Mid-term assessments are a common technique to evaluate learning success [22]. In-class and online-assessments, complement the flip model [51]. Therefore, clear accordance of video tutorials and in-class content (and assessments) is essential [2], [57]. Furthermore, the total student-workload should be considered [58]. The teacher must permanently control and steer the course regarding the students' needs, the planned results and the amount of work.

The kind and tone of interaction are important for the satisfaction emphasized by the learners [50]. The original FC model was designed as a flipped mastery model with little peer interaction and a focus on individual learning [59]. The role of the teacher changes in an FC environment [1], [45]. Recently, the share of team-based learnings and high group interaction increased [37], [59]. In team-based learning settings, the material for the in-class courses is processed iteratively. First, the individual student works with the material; then the individual results are discussed within the groups and finally debated with the teacher and presented in class [37] [59]. These iterations seem quite time-consuming to the students and can lead to a resistance to change [37]. In the worst case this results in absence from class. Some students also regard the in-class courses as obsolete, as they can learn the basic content of the class online. However, as intended by socio-constructivists, the group based learning in FC courses is essential for learning success [60].

In many studies learning success is directly linked to the exams. An FC design does not inevitably lead to a change in the way exams were carried out [61], but the high interaction and the available technical infrastructure enable changes [22]. Exams focused on problem-solving [62] or including bonus points [58] occur. In many cases exams mark the reaching of this milestone. The execution phase ends with *the graduation of the course (M6)*. This milestone marks the end of the interaction between the teacher and the students.

4.4 Closing Phase

The final phase of the FC process model is the closing phase. This is the time to evaluate the course, collect perceptions about the FC construct, content and overall implementation. The closing phase includes the analysis of data obtained on results and perceptions. The analysis is based on the teacher's experience, exam results and measurements of the students' attitude towards the concept [58]. Therefore, differentiation between summative (to measure the outcome) and formative (to formulate the lessons learned and re-design the concept) evaluations is necessary [39].

Most evaluations in FC research are based on self-reported scales using quantitative and qualitative data [22]. These scales often comprise perceptions of feelings, subjective experiences and satisfaction [60]. To assess the attitude towards the FC even scales, close to self-efficacy rate, are taken. We identified scales measuring enjoyment, self-confidence and perceived value of the content [60]. Furthermore, the evaluation of the learning success [11] and the students' effectiveness are important [31]. Many researchers claim the increase of the learning success using an FC scenario [41]. Often the students' performance as a whole increase in FC classes compared to traditional lectures [63]. Kerr mentions that even the middle and the lower third of the examination group increase their performance [20]. These results mainly aim at the exam-outcomes. Further positive statements regarding the problem-solving ability of students exist [7]. Only a few studies show opposite results [30]. Despite the measurable learning success, Foldnes [59] shows, that the increase of group interaction positively influences the learning outcomes. The great variety of evaluation designs shows that there is no standard tool for assessing FC neither formatively nor in a summative way. The evaluation can be further supplemented by LA, providing deeper insights into student's interaction and behavior throughout the whole course [26]. Only a few articles can be found that give room for the *lessons learned (M7)* to develop a sustainable culture of FC classes [9]. Teachers should use the closing phase to reconsider the contents and continuously work on the renewal of the contents and methods [11]. The pre-recorded online material should be critically revised [11]. In FC proceeding, the time made available for post-processing is limited. Nevertheless, it is important to collect thoughts, write down lessons learned and restructure future classes for sustainable success. These steps can improve the FC by design based approaches [9], [30]. This phase is finished with the *assessments and evaluations (M8)* which are held and analyzed. Results are shared with the organization and used for formative and summative re-organization of the course.

5 Discussion and Conclusion

For the deduction of the conceptual process model we chose a project management model based on four phases. We aligned the activities to initialize and implement an FC class. The structure helps to remember all duties and can be used as a checklist (see Figure 4) [12]. Figure 3, combined with the following chapters and the checklist, will give in-depth information about the activities performed, the important milestones and detailed description including further text references for a clear understanding of how to design and run an FC. Within the project initiation phase, both lecturer and institution decide about the introduction of an FC. During the planning phase, the rough and fine concept of the FC are developed. Besides the execution, the planning phase is the most demanding phase. In the execution phase the interaction with the students and the supply of online-material begins. When all in-class and out-of-class activities are finished, the evaluation phase starts. The activities, already described in more detail in chapter 4, are summarized in a checklist in Figure 4. The list is to be understood both as an overview and as a notepad so that all important activities are taken into account.

It is crucial to take enough time to set up an FC. The planning, content conception and coordination between online and attendance phases as well as the intermediate examinations and quizzes require a high level of professional competence. This is why the training of the teachers is so important. Surprisingly, little is reported about this in the literature found. Furthermore, there is a clear need for more conceptual models. Actual findings are dominated by anecdotal articles and presentations of cases [30]. Most articles imply somehow all phases but focus on different aspects. Future research could concentrate on single phases and the tasks or on single tasks covering all phases.

Despite our merits, the research is not free of restrictions and limitations. Review articles built the dominating part of our literature base. We cannot rule out the possibility that there may be some articles dealing with a more specific issue that we have only been able to address in a marginal way. Also, our model has to be evaluated. Regarding the results from our study, we acknowledge that FC is more for students than for pupils, as the learners need a self-paced learning experience. However, FC is not limited to any specified class of students [7].

For the evaluation of the model we plan to set up an FC for students based on the process model in order to gain further insights from our experience with the application of the model and to further concretize the individual steps. Also, we intend to interview FC teachers and students about our model and checklist in order to review our results and expand the model. A large and diverse set of data relating to the project management-centered design of an FC could provide interesting information that would allow the model to be evaluated and gains and barriers to be compared under different conditions and for different stakeholders.

Initiation

- ☐ Define needs and goals
- ☐ Estimate time, staff, and financial expenses
- ☐ Identify stakeholders (lecturers, students, institution)
- ☐ Weigh benefits and costs and make a decision whether FC is useful

Planning

- ☐ Activate project team
- ☐ Train teachers
- ☐ Check if changes/plans are in accordance with university requirements
- ☐ Identify group of learners (context, diversity)
- ☐ Define learning outcomes and levels with the help of taxonomies
- ☐ Inform students in advance
- ☐ Prepare and produce material (choose media type)
- ☐ Choose in-class activities and prepare material
- ☐ Tune in-class and online courses
- ☐ Prepare Learning Analytics
- ☐ Check and provide resources (rooms, technical equipment, LMS)

Execution

- ☐ Provide online material
- ☐ Continuous learning assessment
- ☐ Proceed in-class activities
- ☐ Monitor participants and learning success
- ☐ Steer according to students' needs

Closing

- ☐ Carry out exams
- ☐ Conduct formative evaluation (attitudes, perceptions) of stakeholders
- ☐ Conduct summative evaluation (learning success) of students
- ☐ Lessons learned

Figure 4. Checklist: Development of an FC

References

1. Bergmann, J., Sams, A.: Flip your Classroom: Reach Every Student in Every Class Every Day. International Society for Technology in Education, Alexandria (2012)
2. Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., Peterson, L.: Higher Education Dominance and Siloed Knowledge: A Systematic Review of Flipped Classroom Research. International Journal of Educational Technology in Higher Education 15, 1 (2018)
3. Bishop, J., Verleger, M.A.: The Flipped Classroom: A Survey of the Research. In: ASEE National Conference Proceedings, pp. 1–18. Atlanta, Georgia (2013)

4. Lage, M.J., Platt, G.J., Treglia, M.: Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. The Journal of Economic Education 31, 30–43 (2000)
5. Zuber, W.J.: The Flipped Classroom, a Review of the Literature. Industrial and Commercial Training 48, 97–103 (2016)
6. Gillette, C., Rudolph, M., Kimble, C., Rockich-Winston, N., Smith, L., Broedel-Zaugg, K.: A Systematic Review and Meta Analysis of Student Pharmacist Outcomes Comparing Flipped Classroom and Lecture. American Journal of Pharmaceutical Education 82, 5 (2018)
7. Rahman, A.A., Aris, B., Mohamed, H., Zaid, N.M.: The Influences of Flipped Classroom: A Meta Analysis. In: 6th IEEE Conference on Engineering Education (ICEED), pp. 24–28. IEEE Press, New York (2014)
8. Betihavas, V., Bridgman, H., Kornhaber, R., Cross, M.: The Evidence for 'Flipping out': A Systematic Review of the Flipped Classroom in Nursing Education. Nurse Education Today 38, 15–21 (2016)
9. Hew, K.F., Lo, C.K.: Flipped Classroom Improves Student Learning in Health Professions Education: A Meta-Analysis. BMC Medical Education 18, 38 (2018)
10. Song, Y., Jong, M.S.Y., Chang, M., Chen, W.: Guest editorial: "HOW" to Design, Implement and Evaluate the Flipped Classroom? A Synthesis. Educational Technology & Society 20, 180–183 (2017)
11. O'Flaherty, J., Phillips, C.: The Use of Flipped Classrooms in Higher Education: A Scoping Review. The Internet and Higher Education 25, 85–95 (2015)
12. Project Management Institute: A Guide to the Project Management Body of Knowledge. (2001)
13. Esslinger-Hinz, I., Giovannini, N., Hannig, J.: Der ausführliche Unterrichtsentwurf. Mit Online-Materialien. Beltz, Weinheim, Basel (2013)
14. Helms, R.W., Banefeld, R., Dalpiaz, F.: A Method for the Design of Gamified Trainings. In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS) (2015)
15. Wang, Z.: Research on Teaching Design and Application of Flipped Classroom Mode. In: 2nd International Conference on Education Technology and Information System, pp. 379–383, Atlantis Press (2014)
16. Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts: Working Papers on Information Systems 10 (2010)
17. Smith, V., Devane, D., Begley, C.M., Clarke, M.: Methodology in Conducting a Systematic Review of Systematic Reviews of Healthcare Interventions. BMC Medical Research Methodology 11, 15 (2011)
18. King, W.R., He, J.: Understanding the role and methods of meta-analysis in IS research. Communications of the Association for Information Systems 16, 32 (2005)
19. Lin, H.-C., Hwang, G.-J.: Research Trends of Flipped Classroom Studies for Medical Courses: A Review of Journal Publications from 2008 to 2017 Based on the Technology-Enhanced Learning Model. Interactive Learning Environments 110, 1–17 (2018)
20. Kerr, B.: The Flipped Classroom in Engineering Education: A Survey of the Research. In: 2015 International Conference on Interactive Collaborative Learning (ICL), pp. 815–818. IEEE, Florence (2015)
21. Giannakos, M.N., Krogstie, J., Chrisochoides, N.: Reviewing the Flipped Classroom Research: Reflections for Computer Science Education. In: Proceedings of the Computer Science Education Research Conference, pp. 23–29. ACM, New York (2014)
22. Velegol, S.B., Zappe, S.E., Mahoney, E.: The Evolution of a Flipped Classroom: Evidence-Based Recommendations. Advances in Engineering Education 4, 1–37 (2015)
23. Tan, C., Yue, W-G., Fu, Y.: Effectiveness of Flipped Classrooms in Nursing Education: Systematic Review and Meta-Analysis. Chinese Nursing Research, 192–200 (2017)

24. Said, M.N.H.M., Zainal, R.: A Review of Impacts and Challenges of Flipped-Mastery Classroom. Advanced Science Letters 23, 7763–7766 (2017)
25. Conole, G., Gašević, D., Long, P., Siemens, G.: Message from the LAK 2011 General & Program Chairs. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, vol. 1 (2011)
26. Jovanović, J., Gašević, D., Dawson, S., Pardo, A., Mirriahi, N.: Learning Analytics to Unveil Learning Strategies in a Flipped Classroom. The Internet and Higher Education 33, 74–85 (2017)
27. Lucke, T.: Using Learning Analytics to Evaluate the Effectiveness of the Flipped Classroom Approach. In: Proceedings of the 25th Annual Conference of the Australasian Association for Engineering Education, pp. 1156–1164. Massey University, Wellington (2014)
28. Ward, M., Knowlton, M.C., Laney, C.W.: The Flip Side of Traditional Nursing Education: A Literature Review. Nurse Education in Practice 29, 163–171 (2018)
29. Hu, R., Gao, H., Ye, Y., Ni, Z., Jiang, N., Jiang, X.: Effectiveness of Flipped Classrooms in Chinese Baccalaureate Nursing education: A Meta-Analysis of Randomized Controlled Trials. International journal of nursing studies 79, 94–103 (2018)
30. Lo, C.K., Hew, K.F.: A Critical Review of Flipped Classroom Challenges in K-12 Education: Possible Solutions and Recommendations for Future Research. Research and Practice in Technology Enhanced Learning 12, 4 (2017)
31. Chen, F., Lui, A.M., Martinelli, S.M.: A Systematic Review of the Effectiveness of Flipped Classrooms in Medical Education. Medical Education 51, 585–597 (2017)
32. Shnai, I.: Systematic Review of Challenges and Gaps in Flipped Classroom Implementation: Toward Future Model Enhancement. Proceedings of the European Conference on e-Learning, 484-490, (2017)
33. Mason, G.S., Shuman, T.R., Cook, K.E.: Comparing the Effectiveness of an Inverted Classroom to a Traditional Classroom in an Upper-Division Engineering Course. IEEE Transactions on Education 56, 430–435 (2013)
34. Schöbel, S., Söllner, M.: How to Gamify Information Systems-Adapting Gamification to Individual Preferences. In: Proceedings of the European Conference on Information Systems (ECIS) (2016)
35. Enfield, J.: Looking at the Impact of the Flipped Classroom Model of Instruction on Undergraduate Multimedia Students at CSUN. TechTrends 57, 14–27 (2013)
36. Tucker, B.: The Flipped Classroom - online unstruction at home frees class time for learning. Education Next, 82–83 (2012)
37. Balan, P., Clark, M., Restall, G.: Preparing Students for Flipped or Team-Based Learning Methods. Education and Training 57, 639–657 (2015)
38. Fulton, K.: The Flipped Classroom: Transforming Education at Byron High School: A Minnesota High School with Severe Budget Constraints Enlisted YouTube in Its Successful Effort to Boost Math Competency Scores. Technological Horizons In Education 38, 3 (2012)
39. Vogelsang, K., Liere-Netheler, K., Hoppe, U., Hagerer, I.: Analysis of the Use of Digital Media to Design a Blended Learning Environment by the Example of a Master Course Lecture. In: Proceedings of the 5th International Conference on Education and Information Systems, Technologies and Applications (EISTA) (2017)
40. Anderson, L.W., Krathwohl, P.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition. White Plains, New York (2001)
41. McLean, S., Attardi, S.M., Faden, L., Goldszmidt, M.: Flipped Classrooms and Student Learning: Not Just Surface Gains. Advances in Physiology Education 40, 47–55 (2016)

42. Chen, S.-C., Yang, S.J.H., Hsiao, C.-C.: Exploring Student Perceptions, Learning Outcome and Gender Differences in a Flipped Mathematics Course. British Journal of Educational Technology 47, 1096–1112 (2016)
43. DeLozier, S.J., Rhodes, M.G.: Flipped Classrooms: A Review of Key Ideas and Recommendations for Practice. Educ Psychol Rev 29, 141–151 (2017)
44. Herreid, C.F., Schiller, N.A.: Case Studies and the Flipped Classroom. Journal of College Science Teaching 42, 62–66 (2013)
45. Zappe, S., Leicht, R.: Flipping the Classroom to Explore Active Learning in a Large Undergraduate Course. American Society for engineering education (2009)
46. Grypp, L., Luebeck, J.: Rotating Solids and Flipping Instruction. The Mathematics Teacher 109, 186 (2015)
47. Arnold-Garza, S.: The Flipped Classroom Teaching Model and Its Use for Information Literacy Instruction. Communications in Information Literacy (Comminfolit) 8, 7 (2014)
48. Vazquez, J.J., Chiang, E.P.: Flipping Out! A Case Study on How to Flip the Principles of Economics Classroom. International Advances in Economic Research 21.4, 379–390 (2015)
49. Jensen, J.L., Kummer, T.A., Godoy, D.D.M.: Improvements from a Flipped Classroom May Simply be the Fruits of Active Learning. CBE life sciences education 14 (2015)
50. Findlay-Thompson, S., Mombourquette, P.: Evaluation of a Flipped Classroom in an Undergraduate Business Course. Business Education & Accreditation 6, 63–71 (2014)
51. Demski, J.: Expert Tips for Flipping the Classroom. Campus Technology, 26(5), 32-37 (2013)
52. Milman, N.B.: The Flipped Classroom Strategy: What is it and How Can it Best be Used? Distance Learning 9, 85–87 (2012)
53. Bhagat, K.K., Chang, C. and Chang, C.: The Impact of the Flipped Classroom on Mathematics Concept Learning in High School. Journal of Educational Technology & Society 19, 134–142 (2016)
54. Lehmann, K., Oeste, S., Janson, A., Söllner, M., Leimeister, J.M.: Flipping the Classroom – IT-unterstützte Lerneraktivierung zur Verbesserung des Lernerfolges einer universitären Massenlehrveranstaltung. HMD 52, 81–95 (2015)
55. Danker, B.: Using Flipped Classroom Approach to Explore Deep Learning in Large Classrooms. The IAFOR Journal of Education 3 (2015)
56. Vogelsang, K., Hagerer, I., Hoppe, U., Liere-Netheler, K.: Entwicklung einer Evaluation für Blended Learning Konzepte. Bildungsräume 2017 (2017)
57. Strayer, J.F.: How Learning in an Inverted Classroom Influences Cooperation, Innovation and Task Orientation. Learning Environments Research 15, 171–193 (2012)
58. Vogelsang, K., Hoppe, U.: Development of an Evaluation for Flipped Classroom Courses. Proceeding of Multikonferenz der Wirtschaftsinformatik (mkwi), 821–832 (2018)
59. Foldnes, N.: The Flipped Classroom and Cooperative Learning: Evidence from a Randomised Experiment. Active Learning in Higher Education 17, 39–49 (2016)
60. Foldnes, N.: The Impact of Class Attendance on Student Learning in a Flipped Classroom. Nordic Journal of Digital Literacy 12, 8–18 (2017)
61. Anderson, J., Glenn, H., Frazier, L., Anderson, S.L., Stanton, R., Gillette, C., Broedel-Zaugg, K., Yingling, K.: Comparison of Pharmaceutical Calculations Learning Outcomes Achieved within a Traditional Lecture or Flipped Classroom Andragogy. American Journal of Pharmaceutical Education 81 (2017)
62. Bates, S., Galloway, R.: The Inverted Classroom in a Large Enrolment Introductory Physics Course: A Case Study. In: Proceedings of the HEA STEM Learning and Teaching Conference, 1 (2012)

63. Hibbard, L., Sung, S., Wells, B.: Examining the Effectiveness of a Semi-Self-Paced Flipped Learning Format in a College General Chemistry Sequence. Journal of Chemical Education 93, 24–30 (2016)

The Influence of Risk-Taking on Knowledge Exchange and Combination

Désirée Laubengaier[1], Heinz-Theo Wagner[2], Gerd J. Hahn[1]

[1] German Graduate School of Management & Law, Operations Management and Process Innovation, Heilbronn, Germany
{desiree.laubengaier,gerd.hahn}@ggs.de
[2] German Graduate School of Management & Law, Management and Innovation, Heilbronn, Germany
{heinz-theo.wagner}@ggs.de

Abstract. Knowledge exchange and combination build the core of innovative activity. However, the processes of knowledge exchange and combination are inherently associated with risk. Consequently, for knowledge exchange and combination to occur in an organizational setting, the organization members must be encouraged towards risk-taking. Organizational cultural norms promote a certain behavior among organization members. Accordingly, this study focuses on risk-taking an organizational cultural norm and investigates risk-taking's influence on knowledge exchange and combination.

Keywords: Organizational culture, risk-taking, innovation, knowledge exchange and combination, knowledge management.

1 Introduction

In organization science, knowledge has gained great attention and a "legitimate and important role" [1] has been attributed to it [1, 2]. Knowledge is not only an essential asset for its owners but also an important source of sustainable competitive advantage [3, 4]. Specifically knowledge creation "is essential for the success and survival of firms competing in dynamic environments" [5], and can be considered "the precursor of innovation" [6]. From a knowledge-based perspective, innovation can be thought of as the "creation and application of knowledge to create new knowledge regarding novel products and processes" [6]. Knowledge creation involves two underlying processes, namely knowledge exchange and knowledge combination [7, 8]. These two build the core of any innovative activity which is in accordance with the Austrian School of Economics [8]. Knowledge exchange refers to "interchanging knowledge and information" [9] located in the organization and its members. Knowledge combination, on the contrary, depicts establishing connections between unconnected knowledge pieces or developing new types of connections [7].

Yet, innovative activity is inherently associated with risk [10] and represents a complex endeavor marked by uncertainty [11]. Similarly, knowledge exchange and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

combination are related to risk, and for organization members to engage in exchange and combination, they need to be willing to take risks [7]. In this regard, risk-taking – which is defined as a mindset that encourages the members of an organization to be open towards risks [12] – is essential. Accordingly, extant literature deals with risk-taking in the light of knowledge-related subjects [13], and risk-taking appears to be particularly relevant for knowledge exchange and combination [5, 7]. Research commonly states that risk-taking largely depends on organizational context factors (e.g. [14, 15]). A plethora of studies take the organizational climate for risk-taking into account, and analyze its influence on knowledge creation (e.g. [5]). However, risk-taking might "manifest the cultural variables that exist in and dominate" [16] the organization. This is underlined by several organization scholars who point out that risk-taking is attached to organizational culture (e.g. [12, 17]). In line with current research (e.g. [18]), we consider risk-taking as a norm. Norms define appropriate attitudes and behaviors of organization members [19]. As they convey which conduct is expected and appropriate, they guide organization members' behavior [20]. Accordingly, risk-taking norms direct organization members towards being me more apt to take on risks. As norms are proposed to be influential with regards to knowledge exchange and combination, we assume that risk-taking norms may influence knowledge exchange and combination among organization members significantly. However, empirical research on the influence of risk-taking norms on knowledge exchange and combination is scant.

Considering the above research issues, this study analyzes how risk-taking relates to knowledge exchange and combination. Thereby, we address the following research question: *What is the influence of risk-taking on knowledge exchange and combination?* Drawing on social capital theory, we empirically examine risk-taking's influence on both knowledge exchange and knowledge combination.

2 Related Literature

2.1 Knowledge Exchange and Combination

Social capital theory by Nahapiet and Ghoshal [7] posits that social capital can facilitate knowledge creation which involves two underlying processes, namely knowledge exchange and combination.

Knowledge refers to "a high value form of information that is ready to apply to decisions and actions" [21]. The exchange of knowledge describes "interchanging knowledge and information residing in different organizational members and subunits" [9]. Obviously, knowledge can be located at myriad entities, and its exchange not merely involves, but, necessarily requires more than one sole party. Knowledge exchange continuously provides new knowledge which serves as the basis for new combinations. Combination is the process of bringing together "elements previously unconnected or by developing new ways of combining elements previously associated" [7].

Social capital refers to “the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit” [7]. It is an important driver of knowledge exchange and combination [22] and researchers commonly identify three dimensions of social capital [7, 22, 23]: (1) the structural, (2) the cognitive, and (3) the relational dimension. The relational dimension concerns “assets created and leveraged through relationships” [7]. This dimension pertains to resources that bond the members of a social system and transmit influence on their behavior. One form of social capital that is attached to the relational dimension is norms. Norms represent “a degree of consensus in the social system” [7] such as an organization. Norms that are effectively in use impart collectively binding demands on the members’ conduct and can have a significant impact on their knowledge exchange and combination activities [7].

2.2 Risk-Taking as a Cultural Norm

Conform with Yates and Stone [50], we define risk as “the degree of uncertainty and potential loss that may follow from a given behavior or set of behaviors” [17]. People generally dislike risk – and risk ambiguity even more [24]. However, knowledge exchange and combination processes are associated with risk and organization members’ willingness to take risks is critical for these two processes. Risk-taking refers to taking bold actions, and not only involves venturing into the unknown but also dealing with insecurities [25]. Consequently, organizations need to support and encourage risk-taking so that the organization members can cope with risk [26], and consequently engage in knowledge exchange and combination.

Cultural norms represent “a degree of consensus in the social system” [7] such as an organization. Due to the fact that cultural norms are “socially created standards” [18], they are to be grasped as the collectively binding demands and expectations regarding the conduct within the organizations [19]. They give direction to behavior, and if effectively in use, norms have a powerful effect on organization members’ conduct [7]. Similarly, whether and to what extent organization members take the risk and engage in knowledge exchange and combination could be significantly affected by risk-taking norms. Accordingly, we assume that risk-taking positively affects both knowledge exchange and knowledge combination.

3 Research Method

A survey is used for the means of data collection. All variables are measured with multiple-item scales using a seven-point-Likert-scale ranging from “strongly disagree” to “strongly agree”. Risk-taking, our independent variable, is measured with seven items adapted from Tellis et al. (2009) [26], Hogan and Coote (2014) [27], and Miller and Friesen (1982) [28]. Knowledge exchange (first dependent variable) is measured with 4 items and knowledge combination (second dependent variable) with 5 different items which are all adapted from Shu et al. (2009) [9]. The questionnaire undergoes a preliminary test. Researchers as well as innovation-related employees from different sub-divisions of the organizational unit under investigation serve as a pilot for the survey, helping to ensure that the items are appropriate and clearly worded for this particular sample. Data source is an organizational unit of production control and logistics in a major German, and internationally operating automotive manufacturer. Informants range from top management to lower employees. This sample is appropriate because the organization unit and informants are concerned with knowledge-intense work. We seek to reach an overall sample size of approximately 400 participants. The data collection process consists of two phases. First, emails with the link to the online-questionnaire on the academic survey platform "Unipark" are sent out. In the second step, a reminder-email is sent out after two weeks.is sent out via email.

This study uses structural equation modelling (SEM) for estimating the research model. SEM is appropriate because it allows us to test the relationships between the variables simultaneously.

4 Expected Contribution

We will contribute to extant research by providing empirical insights on risk-taking's influence on both knowledge exchange and knowledge combination. This will shed light on the norm "risk-taking" – which we consider the form of social capital that has the power to shape organization members' behavior to a large extent – and its role for knowledge exchange and combination. Thereby, this study will help create a more thorough understanding of the significance of risk-taking norms for organizational innovation. The results of this study will inform practitioners about the importance of embedding risk-taking as a norm within the organizational culture in order to strengthen organization members' knowledge exchange and combination activities.

References

1. Nonaka, I., Krogh, G. von: Tacit Knowledge and Knowledge Conversion. Controversy and Advancement in Organizational Knowledge Creation Theory. Organization Science 20, 635–652 (2009)
2. Teece, D.J.: Capturing Value from Knowledge Assets. The New Economy, Markets for Know-How, and Intangible Assets. California Management Review 40, 55–79 (1998)
3. Grant, R.M.: Toward a Knowledge-Based Theory of the Firm. Strategic Management Journal 17, 109–122 (1996)
4. Kogut, B., Zander, U.: Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. Organization Science 3, 383–397 (1992)
5. Smith, K.G., Collins, C.J., Clark, K.D.: Existing Knowledge, Knowledge Creation Capability, and the Rate of New Product Introduction in High-Technology Firms. AMJ 48, 346–357 (2005)
6. Shujahat, M., Sousa, M.J., Hussain, S., Nawaz, F., Wang, M., Umer, M.: Translating the Impact of Knowledge Management Processes into Knowledge-Based innovation. The

Neglected and Mediating Role of Knowledge-Worker Productivity. Journal of Business Research (2018)
7. Nahapiet, J., Ghoshal, S.: Social Capital, Intellectual Capital, and the Organizational Advantage. The Academy of Management Review 23, 242 (1998)
8. Schumpeter, J.A.: Theory of Economic Development. Transaction, Cambridge, MA (1934)
9. Shu, C., Page, A.L., Gao, S., Jiang, X.: Managerial Ties and Firm Innovation. Is Knowledge Creation a Missing Link? J Prod Innov Manag 29, 125–143 (2012)
10. Greve, H.R.: A Behavioral Theory of R&D Expenditures and Innovations. Evidence from Shipbuilding. Academy of Management Journal, 685 (2003)
11. Baer, M.: Putting Creativity to Work. The Implementation of Creative Ideas in Organizations. Academy of Management Journal 55, 1102–1119 (2012)
12. Kitchell, S.: Corporate culture, environmental adaptation, and innovation adoption. A qualitative/quantitative approach. Journal of the Academy of Marketing Science 23, 195–205 (1995)
13. Mueller, J.: The Interactive Relationship of Corporate Culture and Knowledge Management. A Review. Rev Manag Sci 6, 183–201 (2012)
14. Pennings, J.M.E., Smidts, A.: Assessing the Construct Validity of Risk Attitude. Management Science 46, 1337–1348 (2000)
15. Sitkin, S.B., Pablo, A.L.: Reconceptualizing the Determinants of Risk Behavior. The Academy of Management Review 17, 9–38 (1992)
16. Zahra, S.A.: Entrepreneurial Risk Taking in Family Firms. Family Business Review 18, 23–40 (2005)
17. Mullins, J.W., Forlani, D., Walker, J.O.C.: Effects of organizational and decision-maker factors on new product risk taking. J Prod Innov Manag 16, 282–294 (1999)
18. Chatman, J.A., Caldwell, D.F., O'Reilly, C.A., Doerr, B.: Parsing organizational culture. How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. J. Organiz. Behav. 35, 785–808 (2014)
19. O'Reilly, C.A., Chatman, J.A.: Culture as Social Control. Corporations, Cults, and Commitment. RESEARCH IN ORGANIZATIONAL BEHAVIOR 18, 157–200 (1996)
20. Parks, L., Guay, R.P.: Personality, values, and motivation. PERSONALITY AND INDIVIDUAL DIFFERENCES 47, 675–684 (2009)
21. Davenport, T.H., Long, D.W. de, Beers, M.C.: Successful Knowledge Management Projects. Sloan Management Review 39, 43–58 (1998)
22. Kang, S.C., Snell, S.A.: Intellectual Capital Architectures and Ambidextrous Learning. A Framework for Human Resource Management. J Management Studies, 65 (2009)
23. Kang, S.-C., Morris, S.S., Snell, S.A.: Relational Archetypes, Organizational Learning, and Value Creation. Extending the Human Resource Architecture. The Academy of Management Review, 236 (2007)
24. Loch, C.H.: Creativity and Risk Taking Aren't Rational. Behavioral Operations in MOT. Production & Operations Management 26, 591–604 (2017)
25. Rauch, A., Wiklund, J., Lumpkin, G.T., Frese, M.: Entrepreneurial Orientation and Business Performance. An Assessment of Past Research and Suggestions for the Future. Entrepreneurship Theory and Practice 33, 761–787 (2009)
26. Tellis, G.J., Prabhu, J.C., Chandy, R.K.: Radical Innovation across Nations. The Preeminence of Corporate Culture. Journal of Marketing 73, 3–23 (2009)
27. Hogan, S.J., Coote, L.V.: Organizational Culture, Innovation, and Performance. A Test of Schein's Model. Journal of Business Research 67, 1609–1621 (2014)
28. Miller, D., Friesen, P.H.: Innovation in Conservative and Entrepreneurial Firms. Two Models of Strategic Momentum. Strategic Management Journal 3, 1–25 (1982)

Gamified Feedback durch Avatare im Mobile Learning

Tim Schneider, Andreas Janson

Universität Kassel, Fachgebiet Wirtschaftsinformatik, Kassel, Deutschland
{tim.schneider,andreas.janson}@uni-kassel.de

Zusammenfassung. Viele Arbeitsprozesse in der Industrie werden zunehmend digitalisiert. Diese Entwicklung erfordert neue Kompetenzen, welche im Arbeitsprozess durch Ansätze wie mobile Applikationen erworben werden können. Aufgrund der Informationsfülle im Mobile Learning und den Spezifika solcher Endgeräte kann es zu einer kognitiven Überlastung des Lernenden kommenden, wodurch dieser Lerninhalte nicht mehr aufnehmen kann. Um diesen Herausforderungen zu begegnen, schlagen wir den Einsatz von Gamified Feedback durch Avatare vor, welche als Lehrerergänzung die notwendigen Hilfestellungen und Informationen bereitstellen können. Hierzu wird im Kurzbeitrag eine Studie zur Untersuchung der Wirkungsweise von Gamified Feedback auf Cognitive Load von Lernenden im Mobile Learning präsentiert.

Keywords: Gamification, Feedback, Mobile Learning

1 Einleitung

Die zunehmende Digitalisierung der Arbeitsprozesse in der Industrie erfordert neue Kompetenzen und Fähigkeiten seitens der Mitarbeiter [1]. Das McKinsey Global Institute geht davon aus, dass 2030 rund 14% der globalen Arbeitskräfte aufgrund der Digitalisierung eine Weiterbildung benötigen, um den neuen Anforderungen an Arbeitskräfte in Zeiten der Digitalisierung gewachsen zu sein [2]. Diese Entwicklung führt zu einem stärkeren Bedarf nach Maßnahmen der Aus- und Weiterbildung (AuW). Gleichzeitig läutet die Digitalisierung einen Paradigmenwechsel in der AuW ein, hin zu informationsgestützten, innovativen Lernsystemen [3]. Diese Lernsysteme weisen einen hohen Flexibilitätsgrad auf, welches insbesondere bei Mobile Learning (ML) zu einer weiten Anwendungsbreite führt und sich in den starken Wachstumsraten von jährlich 8% wiederspiegelt [4].

Jedoch erfordert diese Flexibilität seitens der Lernenden hohe selbstregulatorische Fähigkeiten, um den Lernprozess selbstständig zu organisieren [5]. Diese selbstständige Organisation kann zu einem Überforderungserleben führen, welches sich in Frust und schlechten Lernergebnissen ausdrückt [5]. Um diesen Umstand zu begegnen, wird der Lernprozess in traditionellen Lernsettings von einer Lehrperson begleitet, welche direkt und unmittelbar in den Lernprozess eingreifen kann [6]. Insbesondere Feedback (FB) kann in diesem Zusammenhang zur Verbesserung des Lernerfolgs eingesetzt werden [7]. In digitalen Umgebungen fehlt dieses Angebot durch eine Lehrperson jedoch oft oder ist zu unspezifisch [6]. Deshalb muss der Einsatz

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

von Feedback in ML auf diese Umgebung angepasst werden. Aus diesem Grund schlagen Zichermann und Cunningham [8] den Einsatz von Gamification Elementen in Kombination mit FB vor. Dieses Gamified Feedback kann die den Lernerfolg von Lernenden verbessern [9]. Wir schlagen deshalb den Einsatz eines Avatares als Gamification Element als virtueller Lehrer vor, welcher die Lernenden gezielt unterstützt und ihren Lernprozess strukturiert [10].

Jedoch ist unklar, wie solche Gamified Feedback Elemente gestaltet werden müssen, um eine möglichst effiziente Interaktion zwischen Avatar und Lernenden zu ermöglichen und damit den Lernerfolg bestmöglich zu erhöhen. Zusätzlich ist unklar, wie diese Elemente im Rahmen von Mobile Learning eingesetzt werden müssen, da bspw. die geringe Displaygröße besondere Anforderung an die Gestaltung der Elemente setzt. Hierbei ist im speziellen die kognitive Last (Cognitive Load – CL) der Lernenden zu beachten [11]. Vor diesem Hintergrund beschäftigt sich diese Arbeit mit der folgenden übergeordneten Forschungsfrage (RQ):

RQ: *Wie muss Gamified Feedback durch Avatare im Mobile Learning gestaltet werden, damit der Lernerfolg von Lernenden gesteigert werden kann?*

Um diese Frage zu beantworten präsentieren wir in diesem Kurzbeitrag ein 2x2 Experiment, welches auf Grundlage einer ML Applikation (MLA) für die AuW den Effekt verschiedener Gamified Feedbacktypen untersucht. Dabei ist die Konzeption der MLA in ein größeres Design Science Research (DSR) Projekt [12] eingebettet.

2 Theoretischer Hintergrund

FB wird sowohl in der Offline als auch in der Online Welt zur Steigerung des Lernerfolgs eingesetzt [13], um u.a. die Verbesserung der selbstregulatorischen Fähigkeiten einer Person zu ermöglichen [14]. Dabei ist FB eine der wichtigsten Mittel zur Verbesserung von Motivation, Erfolg und Effizienz [14]. FB wird in formatives (FB während des Lernprozesses) und summatives (FB nach dem Lernprozess) FB unterteilt [7]. Das FB wird über eine Lehrperson an den Lernenden weitergegeben. Jedoch ist in vielen digitalen Lernplattformen ein unmittelbarer Zugang zu einer Lehrperson nicht verfügbar [15]. Durch die Zeit- und Ortsunabhängigkeit im ML [16] ist es besonders herausfordernd FB und Gamification Elemente so zu gestalten, dass sie gut vom Lernenden aufgenommen werden können und als nützlich angesehen werden. Durch die Übertragung von FB auf den Gamification Ansatz soll die Wirkung von FB verstärkt werden [17]. Dabei stellt Gamification einen Ansatz dar, welcher spieltypische Elemente in einem nicht Spielekontext einsetzt, um die Motivation und die Leistung von Nutzern zu steigern [18]. Gamification Elemente können in Spielmechaniken, -dynamiken und -motive eingeteilt werden [8, 19]. Vor diesem Hintergrund definieren wir Gamified Feedback als die Nutzung von spieltypischen Elementen zur Verstärkung der Wirkung von Feedback.

Die Cognitive Load Theory bietet in diesem Kontext ein Rahmenwerk, wie Instruktionsmaterialien gestaltet werden können [11], um die kognitiven Kapazitäten einer Person nicht zu überlasten. Durch die begrenzten Kapazitäten müssen die FB Materialien in der MLA so gestaltet werden, sodass diese bestmöglich vom kognitiven

System des Lernenden aufgenommen werden können. Deshalb müssen Elemente identifiziert werden, welche für Gamified formatives und summatives FB relevant sind, um eine effiziente Informationsweiterleitung zu ermöglichen.

3 Gestaltungsansatz für die Mobile Lernanwendung

Um die Wirkung von Gamified FB mittels Avatar in einer MLA zu evaluieren, wird im ersten Schritt ein Prototyp im Rahmen eines gestaltungsorientierten Projekts in der industriellen Fertigung entwickelt. In diesem Zuge soll die Wirkung von summativen und formativen Gamified FB auf den Lernenden bzgl. des Lernerfolgs und CL untersucht werden. Abbildung 1 zeigt die prototypische Implementierung der MLA.

Abbildung 1: Prototypische Implementierung der MLA

Der Avatar erscheint, sobald sich der Lernende die MLA betritt. Dieser begleitet ihn durch die Anwendung und gibt ihm nach seinem individuellen Fortschritt FB entweder während (formativ) oder nach dem Arbeitsprozess (summativ). Dabei bietet der Avatare u.a. kurze Lernvideos, Bilder oder text- und mediengestützte Quizelemente an. Zudem regt der Avatar die Probanden dazu an, sich selbstständig Lernziele zu setzen, indem er ihnen Nachfragen zu ihrem Lernprozess stellt. Gleichzeitig bietet der Avatar nur die Informationen an, welche für den Arbeitsprozess wichtig sind, um eine Überforderung der Lernenden zu vermeiden. Um dies zu ermöglichen und die MLA auf den neusten Stand bzgl. des Arbeitsprozesses zu halten, verbindet sich diese mittels einer NFC Schnittstelle mit der Maschine, an welcher die Person momentan arbeitet. Diese Schnittstelle ist die Grundlage für das aufgabenbezogene FB. Dadurch soll gewährleistet werden, dass der Lernende situationsbezogenes und relevantes FB für seinen Arbeitsprozess erhält.

4 Vorgehen und Ausblick

Zur Untersuchung von Gamified Feedback in einer MLA in der industriellen AuW wird ein 2x2 Pre-Post Experiment (summatives X formatives Feedback) mit einer Kontrollgruppe durchgeführt. Dabei wird das Feedback durch einen Avatar an die Lernenden vermittelt. Die Kontrollgruppe erhält kein Gamified Feedback vom Avatar. Dabei wird der Lernerfolg mittels des von Keller [20] entwickelten ARCS-Modells, sowie der CL mittels einer objektiven, indirekten Messmethode [21] untersucht.

Während des Experiments durchlaufen die Lernenden den gleichen Arbeits- und Lernprozess (vgl. Abbildung 2).

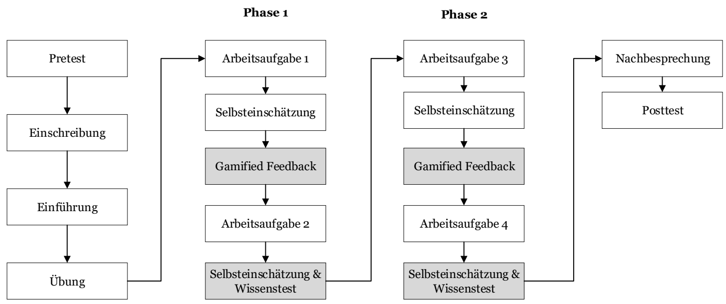


Abbildung 2: Ablauf des Experiments (basierend auf [22])

Die Evaluation wird in Berufsschulen mit industriellem Hintergrund mit Industriemechanikern durchgeführt. In einem Pre-Test werden die Demographika wie Geschlecht oder Vorwissen abgefragt [16]. Zusätzlich wird das Vorwissen bzgl. des Arbeitsprozesses und Erfahrungen bzgl. FB aufgenommen [23]. Darüber hinaus erfassen wir die Kontrollfähigkeit zum selbstorganisierten Lernen [24] .Der CL wird durch eine objektive, indirekte Methodik erfasst [21]. Danach werden die Teilnehmer randomisiert einen der vier Gruppen zugeteilt und müssen vordefinierte Aufgaben lösen. Dabei unterscheiden sich die Gruppen in der Art des Feedbacks. Die Kontrollgruppe erhält kein Gamified FB. Das Feedback wird entweder automatisch während, nach oder proaktiv auf Anforderung des Lernenden angeboten. Der Zeitpunkt der FB Bereitstellung wird hierbei durch die NFC Schnittstelle an die momentane Arbeitssituation angepasst. Dadurch soll sichergestellt werden, dass der Lernende arbeitsprozessbezogenes FB erhält. Für die weitere Vorgehensweise ist ein zweiphasiges Vorgehen geplant. Im ersten Schritt soll die hier vorgestellte experimentelle Analyse erste Aufschlüsse liefern, welche zur Umsetzung eines Feld Experiments in der zweiten Phase genutzt werden können. Durch dieses Vorgehen wollen wir zu Beginn kurzfristige und in der zweiten Phase auch langfristigere Auswirkungen des MLA auf den Lernenden untersuchen.

Basierend auf den von Seaborn und Fels [25] identifizierten Forschungslücken ist der theoretische Beitrag dieser Studie zweigeteilt. Zum einen möchten wir dazu beitragen, dass Verständnis von Gamified FB in ML und deren Auswirkungen auf Lernende besser zu verstehen. Zum anderen wollen wir im Rahmen des DSR Ansatzes präskriptives Wissen für die effektive Gestaltung von Gamified FB in ML gewinnen [26]. Der praktische Beitrag bezieht sich hierbei auf das Gestaltungswissen, welchen MLA Entwicklern helfen soll, wirkungsvolle Anwendungen für die AuW zu designen.

Danksagung

Dieser Artikel wurde im Rahmen des Projekts KoLeArn (FKZ: 01BE17008A) unter der Projektträgerschaft des Deutschen Zentrums für Luft- und Raumfahrt erarbeitet und mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literaturverzeichnis

1. Billett, S., Harteis, C., Gruber, H.: International Handbook of Research in Professional and Practice-based Learning. Springer Netherlands (2014)
2. McKinsey Global Institute: Skill Shift. Automation and the future of the workforce (2018)
3. acatech: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0 (2013)
4. Docebo: Learning on the go: Tips and trends in m-learning - a report (2014)
5. Frohberg, D., Göth, C., Schwabe, G.: Mobile Learning projects - a critical analysis of the state of the art. Journal of Computer Assisted Learning 25, 307–331 (2009)
6. Means, B., Toyama, Y., Murphey, R., Bakia, M., Jonesm Karla: Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies. U.S. Department of Education, 1–94 (2010)
7. Hattie, J., Timperley, H.: The Power of Feedback. Review of Educational Research 77, 81–112 (2007)
8. Schöbel, S., Janson, A.: Is it all about having fun? - Developing a taxonomy to gamify information systems. Twenty-Sixth European Conference on Information Systems (ECIS), 1–20 (2018)
9. Lavouné, È., Monterrat, B., Desmarais, M., George, S.: Adaptive Gamification for Learning Environments. IEEE, 1–13 (2018)
10. Mull, I., Wyss, J., Moon, E., Lee, S.-E.: An exploratory study of using 3D avatars as online salespeople. Jnl of Fashion Mrkting and Mgt 19, 154–168 (2015)
11. van Merriënboer, J.J.G., Sweller, J.: Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. Educ Psychol Rev 17, 147–177 (2005)
12. Thiel de Gafenco, M., Janson, A., Schneider, T.: KoLeArn – Smarte und kontextsensitive Aus- und Weiterbildung für die chinesische Industrie. DeLFI, 1–12 (2018)
13. Su, C.-H., Cheng, C.-H.: A mobile gamification learning system for improving the learning motivation and achievements. Journal of Computer Assisted Learning 31, 268–286 (2015)
14. Armisen, A., Majchrzak, A., Brunswicker, S.: ICIS 2016 Dublin Formative and Summative Feedback in Solutation Generation: The Role of Community and Decision Support System in Open Source Software. Thirty Seventh International Conference on Information Systems (ICIS) (2016)
15. Yi, M.Y., Davis, F.D.: Developing and Validating an Observational Learning Model of Computer Software Training and Skill Acquisition. Information Systems Research, 146–169 (2003)
16. Hamidi, H., Chavoshi, A.: Analysis of the essential factors for the adoption of mobile learning in higher education. A case study of students of the University of Technology. Telematics and Informatics (2017)
17. Zichermann, G., Cunningham, C.: Gamification by design. Implementing game mechanics in web and mobile apps. O'Reilly (2011)
18. Thiebes, S., Lins, S., Basten, D. (eds.): Gamifying Information Systems - A Synthesis of Gamifaction Mechanics and Dynamics (2014)
19. Blohm, I., Leimeister, J.M.: Gamification. Bus Inf Syst Eng 5, 275–278 (2013)
20. Keller, J.M.: Development and Use of the ARCS Model of Instructional Design. Journal of Instructional Development, 2–10 (1987)
21. Korbach, A., Brünken, R., Park, B.: Differentiating Different Types of Cognitive Load: a Comparison of Different Measures. Educ Psychol Rev 30, 503–529 (2018)

22. Santhanam, R., Liu, D., Shen, W.-C.M.: Research Note—Gamification of Technology-Mediated Training: Not All Competitions Are the Same. Information Systems Research 27, 453–465 (2016)
23. Gupta, S., Bostrom, R.: Technology-Mediated Learning: A Comprehensive Theoretical Model. JAIS 10, 686–714 (2009)
24. Zimmerman, B.J.: Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. American Educational Research Journal 45, 166–183 (2008)
25. Seaborn, K., Fels, D.I.: Gamification in theory and action: A survey. International Journal of Human-Computer Studies 74, 14–31 (2015)
26. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly, 611–642 (2006)

Alexa, Can You Help Me Solve That Problem? – Understanding the Value of Smart Personal Assistants as Tutors for Complex Problem Tasks

Rainer Winkler[1], Eva Bittner[2], Matthias Söllner[3,4]

[1] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
rainer.winkler@unisg.ch
[2] University of Hamburg, Department of Informatics, Hamburg, Germany
bittner@informatik.uni-hamburg.de
[3] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
matthias.soellner@unisg.ch
[4] University of Kassel, Information Systems, Kassel, Germany
soellner@uni-kassel.de

Abstract. In recent decades, the number of students per lecturer at universities has constantly risen. In these learning scenarios, individual lecturer support for helping students actively acquiring new knowledge is hardly possible. However, active student behavior is necessary for successful learning. Smart Personal Assistants such as Amazon's Alexa or Google's Home promise to fill this gap by being students' individual tutors. In order to understand what students expect from Smart Personal Assistants as tutors and how they interact with them, we will carry out an experiment. In this research in progress paper, we present our experiment design, where we observe the individual interaction between students and a Smart Personal Assistant tutor and between students and a human tutor applying the same methods in both cases. Drawing on the concepts of parasocial interaction and trust, we derive hypotheses, present the Smart Personal Assistant development and explain the experiment process in detail.

Keywords: computer tutor, Smart Personal Assistant, education, experimental design,

1 Introduction

In recent decades, the number of students per lecturer at universities has constantly risen [1, 2]. In these learning scenarios, lecturers are often forced to present their learning content to the students in a one-directional way with nearly no support for students before and after the lecture. However, the predominant constructivist learning theory states that learning only occurs when students actively acquire new knowledge [3]. This happens best in learning environments where lecturers individually scaffold students' understanding of new learning content [4]. This kind of support is hardly possible due to universities' financial and organizational restrictions. According to the

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

OECD, the number of students at universities rose from 2005 to 2014 by 15 percentage points in the US and 29 in Germany while public spending for education decreased in the same period by 7 percentage points in the US and 1 in Germany [5]. Smart Personal Assistants (SPA) such as Amazon's Alexa or Google's Assistant that are based on current technological developments in artificial intelligence and natural language processing have the potential to fill this gap. SPAs are software programs that communicate with an individual via natural language (voice and text) helping users to perform tasks [6]. Until now, there is no empirical evidence whether and under what conditions students accept SPAs as their personal tutors. Moreover, in past research, SPAs have been mostly used for rather simple and short tasks such as setting an alarm, restaurant reservations, etc. We therefore try to answer the following research questions:

RQ1: How do students accept Smart Personal Assistants as their tutors in complex problem task environments?

RQ2: How does interaction mode (voice vs. text) influence the acceptance of Smart Personal Assistants as tutors in complex problem task environments?

We employ a 2 x 2 (human vs. SPA and voice vs. text) between-subject experiment design with an actual SPA powered with current technology to address these research questions. By this means, our study contributes to literature in the following ways: (1) To the best of our knowledge, this study is amongst the first that empirically investigate the value of SPAs as personal tutors within complex problem task environments. (2) We extend past research of human and computer tutoring by directly comparing four types of tutoring (human text vs. human voice vs. computer text vs. computer voice).

2 Theoretical Background and Hypotheses

2.1 The ICAP-Framework in the Context of Tutoring

The ICAP-Framework is based on the constructivist learning theory. It helps classifying students' cognitive engagement behaviour during learning with four categories: *interactive, constructive, active, or passive* and predicts that they can be ordered by effectiveness *(better learning outcomes)* as **interactive > constructive > active > passive**. A *passive* student behaviour would be attending to the presented instructional information without additional physical activity (e.g. reading a text). An *active* student behaviour includes "doing something physically" (e.g. taking notes). A *constructive* student behaviour requires "producing outputs that contain ideas that go beyond the presented information" (e.g. self-explaining a text). An *interactive* behaviour requires "dialoguing extensively on the same topic, and not ignoring a partner's contribution [7]. One-on-one tutoring with human tutors mostly elicit interactive student behaviour including co-construction and other collaborative spoken activities leading to an increased understanding of learning content and an increased learning process satisfaction [8]. This way of interaction is currently beyond the state of the art in computer tutoring. Nevertheless, recent developments in artificial intelligence, natural language processing and machine learning let us believe that Smart Personal Assistants are able to replace human tutors to some extent. We therefore propose:

H1. There is no difference in acceptance levels between Smart Personal Assistants and human tutors.

2.2 Voice-based vs. Text-based Smart Personal Assistants

Voice-based discourse is the primary form of human-human communication, hence, computer interfaces that communicate via voice provide a more efficient, meaningful, and naturalistic interaction experience [9]. Within the context of tutoring, there are a lot of theoretical positions supporting that spoken interaction is important in tutorial dialogue, compared to communication through typing and printing [10, 11]. One reason for that might be that the gap between thought and speech is much less than the gap between thought and writing, because voice-based discourse is the language of the "mother tongue" [12]. However, studies also showed that people feel uncomfortable when talking with SPAs in public [13]. Nevertheless, voice-based interactions have the potential to foster relationship building between students and tutors [14]. We therefore propose:

H2. Voice-based Smart Personal Assistants show increased acceptance levels compared to text-based Smart Personal Assistants.

2.3 The mediated role of Parasocial Interaction and Trust

The concept of parasocial interaction was first introduced by Horton and Wohl [15], ahead of ubiquitous interactive computer technology. When computers such as robots or AI came to have human-like interfaces, researchers have started to use this concept to investigate the existence of emotional relationship (e.g. feelings of friendship) between people and computers [16]. Furthermore, Trust is vital in one's acceptance of technology [17]. Many studies found out that users must have trust in a technology before they show willingness to adopt it [17, 18]. Especially the affective dimension of trust may be very important for tutors' acceptance. It explains the confidence one places in a partner on the basis of feelings generated by the level of care and concern the partner demonstrates [19]. We therefore hypothesize that type of tutor as well as interaction mode can be mediated by these constructs.

H3a. The effect of type of tutor on tutor's acceptance is mediated by parasocial interaction.

H3b. The effect of type of tutor on tutor's acceptance is mediated by trust.

H3c. The effect of interaction mode on tutor's acceptance is mediated by parasocial interaction.

H3d. The effect of interaction mode on tutor's acceptance is mediated by trust.

3 Overall Experiment Procedure and Treatment Design

To test our hypotheses, we run a laboratory experiment where participants have to individually conduct a 30-minute complex problem task with the help of a SPA or a human tutor. Table 1 shows our treatment design. The human tutor can be perceived as an expert in tutoring (gold-standard). We include a control group that has no tutor. The sample consists of graduate and undergraduate students attending a Swiss business school.

Table 1. Treatment Design (randomly assigned groups)

		Type of tutor		
		No tutor (control)	**SPA tutor**	**Human tutor**
Interaction mode	**Voice-based**	T0 (N=35)	T1 (N=35)	T2 (N=35)
	Text-based		T3 (N=35)	T4 (N=35)

4 Task Design and SPA Development

We decided to choose a complex problem task comparable to the types of problems people face every day. We therefore first derived task requirements from complexity theory and problem-based learning theory [20, 21]. By this means, we chose a task that was already tested in a study on the effectiveness of problem-based learning [22]:

Increasing traffic flow has led to a significant increase in the number of accidents at a major intersection in your city. Several deaths have occurred in recent years. What should the city do about it?

We used Amazon's Alexa and its Skill Development Kit 2.0 with nodeJS to develop our voice-based and text-based SPA. The interaction model and the corresponding instructions of Alexa is based on the problem-solving steps of Kim and Hannafin [23]: *(a) problem identification, (b) problem exploration, (c) problem reconstruction, (d) presentation and communication, (e) reflection and negotiation*. We conducted a pretest with twenty students in order to test the functionality of our programmed SPAs and noted their change requests.

5 Measurement, Analysis and Conclusion

This experiment will be measured with the help of a posttest-questionnaire. *Tutors' acceptance* will be measured with the help of the technology acceptance model proposed by Davis et al. [24]. In specific, we will adapt items from perceived usefulness and perceived ease of use. Furthermore, we will measure the construct *parasocial interaction* with Rubin et al.'s [25] PSI (Parasocial Interaction) Scale. The construct *trust* will be measured with items proposed by Lee and Choi [26]. For analysis, we will use traditional t-tests to show mean differences between the treatment groups.

We are currently analyzing the data of the experiment. First results show that audio-based SPAs are significantly better than text-based SPAs in terms of acceptance. Our expected results from the experiment have a twofold contribution. To the best of our knowledge, this study is amongst the first that empirically investigate the value of SPAs as tutors for complex problem tasks. As this kind of task support has the potential to increasingly enter education in future, it is important to understand how students accept SPAs as their tutors. Moreover, we extend past research on human and computer tutoring that mainly focused on one type of tutoring.

References

1. Nicol, D.J., Macfarlane-Dick, D.: Formative assessment and self-regulated learning. A model and seven principles of good feedback practice. Studies in higher education 31, 199–218 (2006)
2. Oeste, S., Lehmann, K., Janson, A., Söllner, M., Leimeister, J.M.: Redesigning University Large Scale Lectures: How To Activate The Learner. Academy of Management Proceedings 2015 (2015)
3. Glasersfeld, E.v.: Constructivism. The concise Corsini encyclopedia of psychology and behavioral science 6, 19–21 (1987)
4. Vygotsky, L.S.: Mind in society. The development of higher mental process. Cambridge, MA: Harvard University Press (1978)
5. OECD: Education at a Glance 2016: OECD Indicators. OECD, Paris (2016)
6. Harish, A., Naveensankar, K.S., Abdullah, M., Kiruthika Devi, K.: VFF - A Framework for linking Virtual Assistants with IoT. An international Journal of Advanced Computer Technology 1, 58–61 (2016)
7. Chi, M.T.H., Wylie, R.: The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educational Psychologist 49, 219–243 (2014)
8. D'Mello, S.K., Graesser, A., King, B.: Toward Spoken Human-Computer Tutorial Dialogues. Human-Computer Interaction 25, 289–323 (2010)
9. Voice Interaction Design. Morgan Kaufmann, San Francisco (2005)
10. Litman, D.J., Rosé, C.P., Forbes-Riley, K., Vanlehn, K., Bhembe, D., Silliman, S.: Spoken versus typed human and computer dialogue tutoring. International Journal of Artificial Intelligence in Education 16, 145–170 (2006)
11. Pon-Barry, H., Clark B, Schultz, K., Bratt, E., Peters, S. (eds.): Advantages of spoken language interaction in dialogue-based intelligent tutoring systems. Springer (2004)
12. Chafe, W.: Integration and involvement in speaking, writing, and oral literature. Spoken and written language: Exploring orality and literacy 35, 35–54 (1982)
13. Easwara Moorthy, A., Vu, K.-P.L.: Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. International Journal of Human-Computer Interaction 31, 307–335 (2015)
14. Kim, Y., Baylor, A.L., Shen, E.: Pedagogical Agents as Learning Companions. The Impact of Agent Emotion and Gender. Journal of Computer Assisted Learning 23, 220–234 (2007)
15. Hartmann, T., Goldhoorn, C.: Horton and Wohl revisited: Exploring viewers' experience of parasocial interaction. Journal of communication 61, 1104–1121 (2011)
16. Han, S., Yang, H.: Understanding adoption of intelligent personal assistants. Industr Mngmnt & Data Systems 118, 618–636 (2018)
17. Parasuraman, R., Wickens, C.D.: Humans: Still vital after all these years of automation. Human factors 50, 511–520 (2008)
18. Ghazizadeh, M., Peng, Y., Lee, J.D., Boyle, L.N.: Augmenting the Technology Acceptance Model with Trust: Commercial Drivers' Attitudes towards Monitoring and Feedback. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 56, 2286–2290 (2012)
19. Johnson-George, C., Swap, W.C.: Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. Journal of personality and Social Psychology 43, 1306 (1982)
20. Hmelo-Silver, C.E.: Problem-based learning: What and how do students learn? Educational Psychology Review 16, 235–266 (2004)
21. Byrne, D.: Complexity theory and the social sciences: An introduction. Routledge (2002)

22. Gallagher, S.A., Stepien, W.J., Rosenthal, H.: The effects of problem-based learning on problem solving. Gifted Child Quarterly 36, 195–200 (1992)
23. Kim, M.C., Hannafin, M.J.: Scaffolding problem solving in technology-enhanced learning environments (TELEs). Bridging research and theory with practice. Computers & Education 56, 403–417 (2011)
24. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: a comparison of two theoretical models. Management Science 35, 982–1003 (1989)
25. Rubin, A.M., Perse, E.M., Powell, R.A.: Loneliness, parasocial interaction, and local television news viewing. Hum Commun Res 12, 155–180 (1985)
26. Lee, S., Choi, J.: Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies 103, 95–105 (2017)

Track 5:

Data Science & Business Analytics

Track Chairs:

Prof. Dr. Natalia Kliewer
Freie Universität Berlin

Dr. Daniel Schnurr
Universität Passau

Prof. Dr. Christoph M. Flath
Universität Würzburg

Matching with Bundle Preferences: Tradeoff between Fairness and Truthfulness

Martin Bichler, Sören Merting[1], and Aykut Uzunoglu[2]

[1] Department of Informatics, Technical University of Munich, Germany
{bichler,soeren.merting}@in.tum.de
[2] Faculty of Business and Economics, University of Augsburg, Germany,
aykut.uzunoglu@wiwi.uni-augsburg.de

Abstract. Course assignment is a widespread problem in education. Often students have preferences for course schedules over the week. First-Come First-Served (FCFS) is the most widely used rule to assign students to courses in practice, but recent research led to alternatives with attractive properties. Bundled Probabilistic Serial (BPS) is a randomized mechanism satisfying ordinal efficiency, envy-freeness, weak strategy-proofness, and polynomial runtime. We report a first application of BPS in a large-scale course assignment application and discuss advantages over FCFS comparing a number of metrics such as the size, the average rank, the profile, and the popularity of the assignments. The exponential number of possible course schedules is a central problem in the implementation of combinatorial assignment mechanisms. We propose a new way to elicit preferences, which limits the number of parameters a student needs to provide. This yields a computationally very effective tool to solve course assignment problems with thousands of students in practice.

Keywords: Course Allocation, Combinatorial Assignment Problem, Randomization, Field Study

1 Introduction

Course assignment is arguably one of the most widespread assignment problems where money cannot be used to allocate scarce resources. Those problems of assigning students to different courses or whole schedules of courses appear at most educational institutions. Matching with preferences has received significant attention in the recent years. While simple first-come first-served (FCFS) rules are still wide-spread, many organizations adopted matching mechanisms such as the deferred acceptance algorithm [1, 2] or course bidding [3, 4] to allocate scarce course seats. Although many course assignment problems are similar to the widely studied school choice problems with students private preferences for one out of many courses, other applications differ significantly. In particular, students are often interested in schedules of courses across the week. Assigning schedules of courses has been referred to as the combinatorial assignment problem (CAP) [5]. Similar problems arise when siblings should be

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

assigned to the same schools in school choice [6], or couples in the context of the hospital residency matching [7]. Overall, the CAP can be seen as a general form of a distributed scheduling problem.

Although there is a huge body of literature on scheduling, the CAP is specific in a number of ways. First, we can only elicit ordinal preferences and no money must exchange hands. Second, students have private preferences over course schedules and we want to have mechanisms that incentivize students to reveal these preferences truthfully. Third, apart from efficiency, fairness of the allocation is an important concern in matching with preferences [8]. Fourth, the allocation of course schedules is a computationally hard (NP-hard) problem and for the problem sizes with hundreds of students, an exact solution might not be tractable.

The need to assign course schedules rather than courses individually became apparent in an application of matching with preferences at the Technical University of Munich (TUM) that we will discuss. In the initial three semesters, there are large courses with hundreds of students (e.g. on linear algebra or algorithms). These courses include a lecture and small tutor groups. Students need to attend one tutor group for three to four courses in each semester and they have timely preferences over course schedules that need to be considered, which makes it a combinatorial assignment problem. These problems are widespread in academia.

A first and seminal approach to address this challenging problem, the *approximate competitive equilibrium from equal incomes mechanism* (A-CEEI), was published by Budish [5]. In A-CEEI students report their complete preferences over schedules of courses, the mechanism assigns a budget of fake money to each student that she can use to purchase packages (or schedules) of courses. Then an optimization-based mechanism computes approximate competitive equilibrium prices, and the student is allocated her most preferred bundle given the preferences, budgets, and prices. A-CEEI is relaxing design goals such as strategy-proofness and envy-freeness to approximate notions, which makes it a remarkable and practical contribution to a fundamentally hard problem. The mechanism has been shown to be approximately strategy-proof, approximately envy-free, and Pareto efficient. Budish, Cachon, Kessler and Othman [9] reports the empirical results at the Wharton School of Business. In addition, Budish and Kessler [10] summarize the results of lab experiments.

The work was breaking new ground, but the A-CEEI mechanism is also challenging. First, it is not guaranteed that a price vector and course allocation exists that satisfies all capacity constraints. This is not surprising given that prices are linear and anonymous. Second, the problem of computing the allocation problem in A-CEEI is PPAD-complete and the algorithms proposed might not scale to larger problem sizes required in the field [11]. Third, students might not be able to rank-order an exponential set of bundles, which is a well-known problem (aka. missing bids problem) in the literature on combinatorial auctions (with money) [12-14]. The latter is a general problem in CAP not restricted to A-CEEI, which we will discuss in much more detail below.

Randomization can be a powerful tool in the design of algorithms, but also in the design of economic mechanisms. Nguyen, Peivandi and Vohra [15] recently provided two randomized mechanisms for one-sided matching problems, one with cardinal and one with ordinal preferences for bundles of objects. The mechanism for ordinal preferences is a generalization of probabilistic serial [16] called Bundled Probabilistic Serial (BPS). Nguyen, Peivandi and Vohra [15] show that this randomized mechanism is ordinally efficient, envy-free, and weakly strategy-proof. These appealing properties come at the expense of feasibility, but the constraint violations are limited by the size of the packages. In course assignment problems the size of the packages is typically small (e.g., packages with three to four tutor groups) compared to the capacity of the courses or tutor groups (around 30 seats or more). There is no need for prices or budgets, and computationally the mechanism runs in polynomial time, which is important for large instances of the course allocation problem that can frequently be found. This makes BPS a practical approach to many problems that appear in practice.

1.1 Contribution

We report on a field study and address issues in the implementation of mechanisms for the combinatorial assignment problem that are beyond a purely theoretical treatment. In particular, preference elicitation is a central concern in combinatorial mechanisms with a fully expressive bid language. Theoretical contributions of assignment mechanisms largely focus on envy-freeness and efficiency as primary design desiderata. We also report on properties such as their size, their average rank, the probability of matching, the profile, and the popularity. These properties often matter in the choice of mechanisms beyond traditional ways to measure fairness and efficiency. For market designers it is important to understand the trade-offs.

Overall, we report on the assignment of 1415 students in the summer term 2017 to 67 tutor groups for four classes and the assignment of 1736 students in the winter term 2017/2018 to 66 tutor groups for four classes at the TUM using BPS.[1] For such a large application, we could not elicit preferences of students for BPS and let them participate in FCFS simultaneously. Instead, we simulated FCFS via a version of Random Serial Dictatorship that allows for bundles (BRSD), which is of independent interest as an assignment mechanism.

Finally, we contribute an approach that is applicable in a wide array of CAP applications where timely preferences matter. We elicit a small number of parameters about breaks and preferred times and days of the week. Together with some prior knowledge about student preferences, this allows us to score and rank-order all possible packages

2 The Combinatorial Assignment Problem

Let us now define the combinatorial assignment problem (CAP) in the context of course assignment applications, desirable properties, and randomized mechanisms.

---

[1] Not all students submitted a non-empty preference list. Therefore, we consider in our evaluation not all of the participating students (1439 in summer term, 1778 in winter term).

2.1 The Problem

Assigning objects to agents with preferences but without money is a fundamental problem referred to as *assignment problem* or *one-sided matching with preferences*. We will use the term assignment or matching interchangeably. In course assignment, students express ordinal preferences, which need to be considered in the assignment. A *one-sided one-to-many course assignment problem* consists of a finite set of n students (or agents) S and a finite set of m courses (or objects) C with the *maximum capacities* $q = (q_1, q_2, \ldots, q_m)$.

In the *combinatorial assignment problem* in the context of course allocation, every student $i \in S$ has a complete and transitive preference relation over subsets (or bundles) of elements of C. A preference profile $\succcurlyeq \in \wp^{|S|}$ is an n-tuple of preference relations.

We can model the demand of the students with binary vectors $b \in \{0,1\}^m$, where $b_j = 1$ if course j is included in b. We define the size of a bundle b with $size(b) = \sum_{j=1}^{m} b_j$, the number of different courses included in the bundle. Let B be the set of all feasible bundles b. Let $x_{ib} \in \{0,1\}$ be a binary variable describing if bundle b is assigned to student i. Then we can model the demand and supply as linear constraints. The supply constraints make sure that the capacity of the courses are not exceeded, and the demand constraints determine that each student can win at most one bundle.

$$\sum_{i \in I, b \in B} x_{ib} b_j \leq q_j, \qquad \forall \text{ j} \in \text{C}, \qquad (Supply)$$

$$\sum_{b \in B} x_{ib} \leq 1, \qquad \forall\, i \in S, \qquad (Demand)$$

Courses in our application are actually tutor groups and each tutor group belongs to one of ℓ classes. Students in our application can only select bundles with at most one tutor group in each of these classes. As a result, the possible size of a bundle b is $size(b) \leq \ell \ll m$.

A *deterministic combinatorial assignment* (deterministic matching) is a mapping $M \subset S \times B$ of students S and bundles B of courses C. $\mathcal{M}$ describes the set of all deterministic matchings. A matching is *feasible* if it is a feasible integer solution to the constraints (Demand) and (Supply).

Random combinatorial assignments (random matchings) are related to fractional assignments with $0 \leq x_{ib} \leq 1$ and random assignment mechanisms can be used to fractionally allocate bundles of course seats to students.

For (non-combinatorial) assignment problems with single-unit demands the Birkhoff-von-Neumann theorem [17, 18] says that any random assignment can be implemented as a lottery over feasible deterministic assignments, such that the expected outcome of this lottery equals the random assignment. However, the Birkhoff-von-Neumann theorem fails when students submit preferences for bundles of course seats. Nguyen, Peivandi and Vohra [15] generalize this result and show that any fractional solution respecting the (Demand) and (Supply) constraints can be implemented as a lottery over integral allocations that violate the (Supply) constraints only by at most $\ell - 1$ course seats.

2.2 First-Order Design Desiderata

Efficiency, envy-freeness, and strategy-proofness are design desiderata of first-order importance typically considered in the theoretical literature on deterministic assignment problems. For randomized mechanisms one has to reconsider these design desiderata and we will briefly introduce relevant definitions in this section. Stochastic dominance (SD) is the key concept among all of these definitions as it provides a natural way to compare random assignments. Let Δ describe the set of all possible random matchings. With p_i we refer to the assignment of student i in the random matching p, and denote with p_{ib} the probability that student i gets allocated bundle b. We will omit the subscript i when it is clear which student is meant. Given two random assignments $p, q \in \Delta$, student i SD-prefers p to q if, for every bundle b, the probability that p yields a bundle at least as good as b is at least as large as the probability that q yields a bundle at least as good as b. More formally, a student $i \in S$ SD-prefers an assignment $p \in \Delta$ over $q \in \Delta$, $p \succcurlyeq_i^{SD} q$, if $\sum_{b' \succcurlyeq_i b} p_{ib'} \geq \sum_{b' \succcurlyeq_i b} q_{ib'}, \forall b \in B$. In other words, a student i prefers the random assignment p to the random assignment q if p_i stochastically dominates q_i. Note, that $\succcurlyeq^{SD}$ is not a complete relation. That is there might be assignments p and q, which are not comparable with this relation.

One desirable property of matchings is *(Pareto) efficiency* such that no student can be made better off without making any other student worse off. That is, a random assignment $p \in \Delta$ is *ex post efficient*, if p can be implemented into a lottery over Pareto efficient deterministic assignments. A random assignment $p \in \Delta$ is *ordinally efficient*, if there exists no random assignment q such that q stochastically dominates p, i.e. $\nexists q \in \Delta: \forall i \in S: q \succcurlyeq_i^{SD} p$ and $\exists i \in S: q \succ_i^{SD} p$. Ordinal efficiency comes from the Pareto ordering induced by the stochastic dominance relations of individual students. It can be shown that ordinal efficiency implies ex post efficiency [16].

Fairness is another important design goal. A basic notion of fairness for randomized assignments is the *equal treatment of equals*, i.e. students with identical preferences receive identical (symmetric) random allocations. Envy-freeness is stronger. A random assignment $p \in \Delta$ is *(strongly) SD-envy-free*, if $\forall i, j \in S: p_i \succcurlyeq_i^{SD} p_j$. We call p *weakly SD-envy-free*, if $\nexists i, j \in S: p_j \succ_i^{SD} p_i$. SD-envy-freeness means that student i weakly SD-prefers the random matching she is faced with to the random assignment offered to any other student, i.e., a student's allocation stochastically dominates the outcome of every other student. For weak SD-envy freeness it is only demanded that no student's allocation is stochastically dominated by the allocation of another student. SD-envy-freeness implies equal treatment of equals.

A *randomized assignment mechanism* is a function $\psi: \mathcal{P}^{|S|} \rightarrow \Delta$ that returns a random matching $p \in \Delta$. The mechanism $\psi(\succcurlyeq) = p$ is *ordinally efficient* if it produces ordinally efficient allocations. In terms of fairness, one could aim for a matching where equals are treated equally. We call a randomized matching mechanism *ψ symmetric*, if for every pair of students i and j with $\succcurlyeq_i = \succcurlyeq_j$ also $p_i = p_j$. This means that students who have the same preference profile also have the same outcome in expectation. A randomized mechanism is *envy-free* if it always selects an envy-free matching.

An important property of a mechanism is *strategy-proofness*. This means, that there is no incentive for any student not to submit his truthful preferences, no matter which preferences the other students report. A random assignment mechanism is (strongly) SD-strategy-proof if for every preference profile $\succcurlyeq$, and for all $i \in S$ and $\succcurlyeq_i'$ we have $\psi(\succcurlyeq_i, \succcurlyeq_{-i}) \succcurlyeq_i^{SD} \psi(\succcurlyeq_i', \succcurlyeq_{-i})$.

A random assignment rule ψ is *weakly* SD-strategy-proof if for every preference profile $\succcurlyeq$, there exists no $\succcurlyeq'$ *for some* student $i \in S$ such that $\psi(\succcurlyeq_i', \succcurlyeq_{-i}) \succcurlyeq_i^{SD} \psi(\succcurlyeq_i, \succcurlyeq_{-i})$. We will omit the prefix SD for brevity in the following. It has been shown that participants in strategy-proof mechanisms such as the Vickrey auction do not necessarily bid truthfully in practice. Therefore, there was a recent discussion about obvious strategy-proofness of extensive form games [19].

Definition 1: OSP [19]. A strategy σ is obviously dominant if, for all other strategies σ', at any earliest information set where σ and σ' diverge, the best possible outcome from σ' is no better than the worst possible outcome from σ. A mechanism is obviously strategy-proof (OSP) if it has an equilibrium in obviously dominant strategies.

In section 4.1 we introduce a number of additional design goals that often matter in the practice and that we analyze empirically.

2.3 Mechanisms

A lot is known about assignment problems with single-unit demand. Random Serial Dictatorship (RSD) selects a permutation of the agents uniformly at random and then sequentially allows agents to pick their favorite course among the remaining ones. Gibbard [20] showed that random dictatorship is the only anonymous and symmetric, strongly SD-strategy-proof, and ex post efficient assignment rule when preferences are strict. Pycia and Troyan [21] prove that RSD is a unique mechanism that is obviously strategy-proof, efficient, and symmetric in mechanisms without transfers. However, RSD is not always ordinally efficient, only ex post efficient [16]. Zhou [22] actually showed that no random mechanism for assigning objects to agents could satisfy strong notions of strategy-proofness, ordinal efficiency, and symmetry simultaneously with more than three objects and agents. So, we also cannot hope for these properties in combinatorial assignment problems. RSD can also be applied to the combinatorial assignment problem. The Bundled Random Serial Dictatorship (BRSD) orders the students randomly and assigns the most preferred bundle, which is still available to each student in this order. Although the package preferences take some toll on the runtime, it is still very fast.

First-come first-served (FCFS) can be seen as a serial dictatorship. Students login at a certain registration and then reserve the most preferred bundle of courses that is still available. Although the arrival process is not uniform at random, students have little control over who arrives first. While there is a certain time when the registration starts, hundreds of students log in simultaneously to get course seats and it is almost random who arrives first. We will simulate FCFS via BRSD and run the algorithm repeatedly to get estimates for performance metrics of FCFS.

Probabilistic Serial (PS) [16] produces an envy-free assignment with respect to the reported unit-demand preferences, and it is ordinally efficient, but it is only weakly SD-strategy-proof. Bundled Probabilistic Serial (BPS) by Nguyen, Peivandi and Vohra [15] is a generalization of PS to the combinatorial assignment problem and computes a fractional solution to (Demand) and (Supply). The BPS mechanism is also ordinally efficient, envy-free, and weakly strategy-proof if preferences are strict.

Informally, in BPS all agents *eat* their most preferred bundle in the time interval $[0,1]$ simultaneously with the same speed as long as all included objects are available. As soon as one object is exhausted, every bundle containing this object is deleted and the agents continue eating the next available bundle in their preference list. The duration with which every bundle was eaten by an agent specifies the probability for assigning this bundle to this agent. After a fractional solution x^* was found via BPS, it is implemented as a lottery over integral matchings satisfying the (Demand) and the relaxed (Supply) constraints, as described in [15].

3 Preference Elicitation

The Department of Informatics has been using stable matching mechanisms for the assignment of students to courses since 2014 [2]. The system provides a web-based user interface and every semester almost 1500 students are being matched to lab courses or seminars via the deferred acceptance algorithm for two-sided matching or random serial dictatorship for one-sided matching problems.

In the context of the study reported in this paper, the web-based software was extended with BPS, the lottery mechanism for decomposing fractional solutions, and BRSD. During the winter term 2017/2018, 1778 computer science and information systems students in their third semester participated in the matching and could choose bundles of tutor groups out of four classes. A computer science student could have more than 700,000 different bundles.[2]

A naive approach would be to let the students rank bundles on their own by choosing the time slots they want to have in their preference list. This would take a lot of time and lead to a substantial missing bids problem.

Budish, Cachon, Kessler and Othman [9] describe the preference elicitation used at the Wharton School of Business. Students could report cardinal item values on a scale of 1 to 100 for any course they were interested in taking. In addition, they could report adjustments for pairs of courses, which assigned an additional value to schedules that had both course sections together. Afterwards courses were scored and transformed into an ordinal ranking over feasible schedules. The authors argue that they felt that “adding more ways to express non-additive preferences would make the language too complicated”. Wharton also provided a decision support tool listing the 10 most-preferred bundles, which allowed students to inspect top-ranked schedules and modify the cardinal values.

[2] The computer science students need tutorials from all four classes ($< 22 \cdot 25 \cdot 26 \cdot 52$).

However, our course allocation problem has a more special structure such that we can allow preferences that are more complex without asking for different weights for the courses. We developed an algorithm that allows to rank-order all possible packages based on a few parameters that students need to specify. For this, we can leverage prior knowledge about timely preferences of students for schedules of tutorials and lectures.

Students' preferences mainly concern their commute and the possibility to free large contiguous blocks of time (e.g., a day or a half-day) that they can plan for other activities (e.g., a part-time job). In larger cities such as Munich, the time that students spend for commuting is significant. Also long waiting times between courses are perceived as a waste of time as it is often hard for them to work productively in several one- or two-hour breaks without appropriate office facilities available. For example, if a student had a tutorial on linear algebra in the morning, a lunch break, and then the tutorials for algorithms and software engineering in the afternoon of the same day with the minimal time for breaks specified, this would be considered ideal. The desired length for breaks between tutorials and for the lunch break are considered parameters with default values in the preference elicitation.

First, students choose the lectures and tutorials they are interested in. The selected lectures will be considered in the bundle generation as constraints, i.e. if a time slot of a tutorial overlaps with the time of a selected lecture, then it will no longer be considered in order to allow students to participate in the lecture. In a second step, the student marks available time ranges in a *weekly schedule*. The day is partitioned into weekdays and time blocks of 30 minutes from 8:00 AM to 8:30 PM. If a tutorial is selected, all time slots of this tutorial will be highlighted with a specific color. Thus, students learn when the tutorials and lectures of interest take place.

A student can set a minimal amount of time for a lunch break and a minimal amount of time in-between two events (default value is 15 minutes). We also allow students to provide weights $\{1, \ldots, 5\}$ for the different days. That is, the students can express preferences over the days. The main web page and the main steps a student had to take are summarized in Figure 1.

The preferences elicited on this screen are input for an algorithm that uses prior knowledge about student preferences to rank-order all possible packages. The algorithm first generates bundles that satisfy all constraints and then ranks them. Finding the bundles that do not violate constraints (e.g., lectures to be attended) of the students can be cast as a *constraint satisfaction problem*. After the feasible bundles are generated, we rank these bundles. For this, we assign a score to each bundle that considers how many days a student needs to come to the university per week in total, the preference ordering over the days, the total time a student has to stay at the university each day, and the length of the lunch breaks between courses.

On the ranking page, we display the 30 top rated pre-ranked bundles and the students can adapt this ranking manually, go back to the previous screen and adapt the input parameters, or just accept the ranking with a single click (see Figure 2).

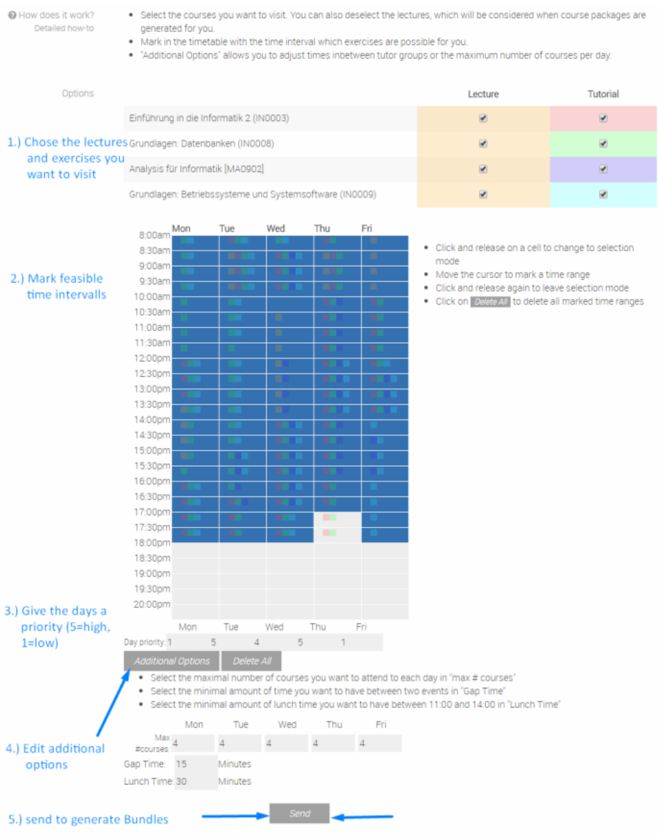


Figure 1. Process to rank-order packages

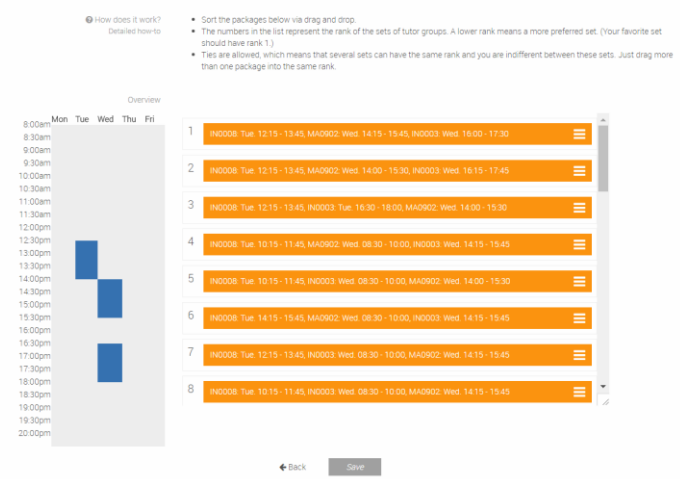


Figure 2. Page with top-ranked packages

Note that $\approx 90\%$ of the students received one of their top ten ranked packages and only a few students received a package with a rank less than 30. So, if a student inspects and confirms the ranking of the first 10-30 packages, this covers the most important quantile of the overall ranking list. We generated a ranking over 200 bundles for each student in advance based on the pre-specified parameters and further preferences only if necessary[3].

4 Results

In Section 2.2 we have summarized first-order design goals for assignment problems: strategy-proofness, fairness, and efficiency. Now we introduce second-order design goals and respective metrics allowing us to compare the assignments of BPS and FCFS empirically and provide numeric results of our two matching instances.

4.1 Metrics

Apart from efficiency, fairness, and strategy-proofness, *popularity* was raised as a design goal. An assignment is called popular if there is no other assignment that is preferred by a majority of the agents. Popular deterministic assignments might not always exist, but popular random assignments exist and can be computed in polynomial time [23]. However, Brandt, Hofbauer and Suderland [24] prove that popularity is incompatible with very weak notions of strategy-proofness and envy-freeness, but it is interesting to understand the popularity of BPS vs. BRSD. In our empirical evaluation, we analyze whether BPS or FCFS are more popular. To measure popularity we first define the function $\phi_i(b,b'): B \times B \to \{\pm 1, 0\}$ associated with the preference relations, where $\phi_i(b,b') = 1$ if $b \succ_i b'$, -1 if $b' \succ_i b$ and 0 in any other case.

Definition 2: Popularity. A random assignment $p \in \Delta$ is *more popular* than an assignment q, denoted $p \rhd q$, if $pop(p,q) > 0$ with $pop(p,q) = \sum_{i \in S} \sum_{b,b' \in B} p_{ib} \cdot q_{ib'} \cdot \phi_i(b,b')$. A random assignment p is *popular*, if $\nexists q \in \Delta: q \rhd p$.

Apart from popularity, the size and the average or median rank are of interest. The *size* of a matching simply describes the number of matched agents. The *average rank* is only meaningful in combination with the size of the matching, because a smaller matching could easily have a smaller average rank. We report the average rank, because it has been used as a metric to gauge the difference in welfare of matching algorithms in [9] and [25], two of the few experimental papers on matching mechanisms.

The *profile* contains more information as it compares how many students were (fractionally) assigned to their first choice, how many to their second choice, and so on.

[3] So far, we described the user interface for the winter term 2017/18. The user interface in the summer term 2017 required students to explicitly drag and drop the pre-ranked packages on a screen. This was considered tedious such that in the winter term the generated ranking was suggested to students right away without any drag-and-drop activities required and could be confirmed without much effort.

The profile of two matchings is not straightforward to compare. We want to compare multiple profiles based on a single metric, and decided to use a metric similar to the *Area under the Curve of a Receiver Operating Characteristic* in signal processing [26], which was already used by [27]. With R denoting the number of possible ranks and $b \in B$, the *Area Under the Profile Curve Ratio (AUPCR)* for matching M can be defined as:

$$AUPCR(M) = \frac{1}{R}\sum_{r=1}^{R} \frac{|\{(i,b) \in M \mid rank(i,b) \leq r\}|}{|S|}$$

4.2 Empirical Results

Due to space constraints, we only analyze the results for the matching in winter term 2017/2018. This application comprised 1736 students and 66 courses (see Table 1). Overall, we had a list of 20,845 different bundles. We simulated FCFS via BRSD on the preferences collected for the BPS. BPS is weakly strategy-proof and in such a large application, it is fair to assume that students do not have sufficient information about the preferences of others, which would be necessary to strategically misreport their preferences. To compare the result of BPS and BRSD we actually would have to run the BRSD for all permutations of the students. Note that computing probabilities of alternatives in RSD explicitly is **#P**-complete [28]. We ran BRSD 1000 to 1,000,000 times with the same preferences but random permutations of the order of students and derived estimates for the different metrics. These estimates are close (see Table 1).

For our data, BPS is more popular than BRSD(1000000). 754 students prefer BPS to FCFS, while 120 students prefer FCFS to BPS (see Table 2). A positive popularity score as described in Definition 2 means, that BPS is more popular than the BRSD outcome and the score for BPS is 3.41 (compared to BRSD(1000000)).

Table 1. Summary statistics for the winter term 2017/2018.

Metric	*BPS*	*BRSD(1000)*	*BRSD(1000000)*
exp rank	1.97372	1.9784	1.97873
exp size	1603.01	1601.03	1600.84
prob match (top 100)	0.923394	0.922253	0.922142
AUPCR	0.889512	0.888184	0.888058
weak envy	0	427	451
strong envy	0	1050	1202

Table 2. Popularity and stochastic dominance of BPS vs. BRSD. The syntax for the SD-preference is the number of students preferring (BPS | BRSD(x)).

Metric	*BRSD(1000)*	*BRSD(1000000)*
popularity winter	1.93061	3.41499
SD-prefer winter	(690|299)	(754|120)

Table 1 reports that for all metrics BPS achieves better results. In the BPS outcome 89.047% of the students receive an assignment ranked in their top ten while in BRSD 88.891% receive such an outcome (see Table 3 for BPS and 4 for BRSD with 1 mio. permultations of the students). The computation times were negligible for BRSD (0.013 seconds per run). BPS required 0.382 seconds, but the lottery algorithm around 30 minutes due to the high number of bundles generated in the winter term.

Table 3. Rank profiles for BPS in winter term 2017/2018.

Rank	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Prob match (%)	73.596	7.083	3.392	1.660	1.041	0.698	0.465	0.447	0.366	0.299
AUPC in (%)	73.596	80.678	84.070	85.730	86.772	87.470	87.935	88.381	88.747	89.047

Table 4. Rank profiles for BRSD(1000000)in winter term 2017/2018.

Rank	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Prob match (%)	73.452	7.046	3.382	1.673	1.040	0.704	0.486	0.443	0.358	0.307
AUPC in (%)	73.452	80.497	83.879	85.553	86.593	87.297	87.783	88.226	88.584	88.891

Our experiments confirm the theoretical result that BPS is (strongly) envy-free. BRSD is neither weakly nor strongly envy-free. In the winter term 1202 students do not SD-prefer their outcome over the outcomes of every other student, and 451 of those students even prefer an outcome of another student (see BRSD(1000000) in Table 1).

4.3 Discussion of Differences

The results from our field experiments and the survey reveal a number of interesting insights. Overall, BPS dominates BRSD on all metrics from our empirical evaluation in both field studies. It has a better average rank, a higher average size and a higher probability of matching, and it does not exhibit envy. However, the differences in average rank, average size, and the profile curve (AUPCR) are small, which is interesting given the fact that only a small number of preferences per student are considered via FCFS.

There are a number of reasons that help to explain the close performance of BPS and FCFS in these metrics. First, Che and Kojima [29] find that random serial dictatorship and probabilistic serial become equivalent when the market becomes large, i.e. the random assignments in these mechanisms converge to each other as the number of copies of each object type grows, and the inefficiency of RSD becomes small. Our empirical results suggest that differences might also be small in large combinatorial assignment markets with limited complementarities.

Second, ordinal preferences do not allow expressing the intensity of preferences. Suppose there are two students who both prefer course c_1 to c_2, each having one course seat only. No matter who gets course c_1, the average rank and size of the matching as well as the profile will be the same even though one student might desperately want to attend c_1, while the second student only has a mild preference for c_1. Without cardinal information about the intensity of a preference, the differences in aggregate metrics can be small.

5 Conclusions

We report two large field studies and show that BPS performs well on a number of additional criteria including average rank, average size, probability of a matching among the first 100 ranks, and the overall profile of ranks (in terms of AUPC of a specific rank) assuming a complete, truthful, and strict ranking of all packages. The matching based on BPS is also more popular than BRSD based on the preferences submitted for BPS. The level of envy in FCFS is significant, even though the size of the packages that can be submitted is limited to the number of classes (three to four groups per package).

The assignment of tutor groups is specific as preferences are mainly about times of the week. The preferred time slots in a week are different from student to student. However, the way how tutor groups should be ordered within these time slots (e.g., time for breaks) can be described with a few parameters such that it was possible to generate packages according to a score.

The paper highlights basic trade-offs in market design without money: FCFS can be seen as a version of serial dictatorship, which is ex post efficient, and obviously strategy-proof and treats students equally. It is also transparent and simple to implement and to understand for students. BPS is a new randomized mechanism that is only weakly strategy-proof, but envy-free, and ordinally efficient. Note that these properties hinge on the availability of strict preferences over all, exponentially many, bundles.

Even if the missing bids problem can be addressed, two important problems remain: First, in contrast to FCFS the BPS mechanism is not obviously strategy-proof[4]. Second, the assumption of strict preferences is strong in the presence of exponentially many bundles. Unfortunately, extending PS or BPS to preferences with ties is not without loss. On the one hand, Katta and Sethuraman [30] extended PS to preferences with indifferences and showed that it is not possible for any mechanism to find an envy-free, ordinally efficient assignment that satisfies even weak strategy-proofness as in the strict preference domain. On the other hand, with indifferences and random tie breaking efficiency cannot be guaranteed. Our preference elicitation technique generates a strict and complete ranking of course bundles based on a few input parameters and is one way to address these issues.

The key difference between BPS and FCFS is the absence of envy. The level of envy in FCFS is significant. Note, that it might be even more pronounced if students were allowed to pick larger packages. If envy-freeness matters, the elegant BPS mechanism has a number of attractive properties and is computationally much less expensive compared to A-CEEI.

[4] Remember that our empirical comparisons are based on the preferences reported in BPS. A part of these preferences might not have reflected the true preferences of participants, and the comparison might be biased towards BPS.

References

1. Gale, D., Shapley, L.S.: College Admissions and the Stability of Marriage. American Mathematical Monthly 69, 9-15 (1962)
2. Diebold, F., Aziz, H., Bichler, M., Matthes, F., Schneider, A.: Course allocation via stable matching. Business & Information Systems Engineering 6, 97-110 (2014)
3. Krishna, A., Ünver, M.U.: Research Note - Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies. Marketing Science 27, 262-282 (2008)
4. Sönmez, T., Ünver, M.U.: Course Bidding at Business Schools. International Economic Review 51, 99-123 (2010)
5. Budish, E.: The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes. Journal of Political Economy 119, 1061-1103 (2011)
6. Abdulkadiroğlu, A., Pathak, P., Roth, A.E., Sönmez, T.: Changing the Boston School Choice Mechanism. National Bureau of Economic Research (2006)
7. Ashlagi, I., Braverman, M., Hassidim, A.: Stability in large matching markets with complementarities. Operations Research 62, 713-732 (2014)
8. Roth, A.E.: The Economics of Matching: Stability and Incentives. Mathematics of Operations Research 7, 617-628 (1982)
9. Budish, E., Cachon, G.P., Kessler, J.B., Othman, A.: Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation. Operations Research 65, 314-336 (2017)
10. Budish, E., Kessler, J.: Can Agents "Report Their Types"? An Experiment that Changed the Course Allocation Mechanism at Wharton. Chicago Booth Working Paper (2017)
11. Othman, A., Papadimitriou, C., Rubinstein, A.: The complexity of fairness through equilibrium. ACM Transactions on Economics and Computation (TEAC) 4, 20 (2016)
12. Milgrom, P.: Simplified Mechanisms with Applications to Sponsored Search and Package Auctions. Games and Economic Behavior 70, 62-70 (2010)
13. Bichler, M., Schneider, S., Guler, K., Sayal, M.: Compact bidding languages and supplier selection for markets with economies of scale and scope. European Journal on Operational Research 214, 67-77 (2011)
14. Bichler, M., Goeree, J., Mayer, S., Shabalin, P.: Simple Auctions For Complex Sales: Bid Languages and Spectrum Auction Design. Telecommunications Policy 38, 613-622 (2014)
15. Nguyen, T., Peivandi, A., Vohra, R.: Assignment Problems with Complementarities. Journal of Economic Theory 165, 209 - 241 (2016)
16. Bogomolnaia, A., Moulin, H.: A new solution to the random assignment problem. Journal of Economic theory 100, 295-328 (2001)
17. Birkhoff, G.: Three observations on linear algebra. Univ. Nac. Tucumán. Revista A 5, 147-151 (1946)
18. Von Neumann, J.: A certain zero-sum two-person game equivalent to the optimal assignment problem. Contributions to the Theory of Games 2, 5-12 (1953)
19. Li, S.: Obviously strategy-proof mechanisms. American Economic Review 107, 3257-3287 (2017)
20. Gibbard, A.: Manipulation of schemes that mix voting with chance. Econometrica: Journal of the Econometric Society 665-681 (1977)
21. Pycia, M., Troyan, P.: Obvious dominance and random priority. Working Paper (2018)
22. Zhou, L.: On a conjecture by gale about one-sided matching problems. Journal of Economic Theory 52, 123-135 (1990)
23. Kavitha, T., Mestre, J., Nasre, M.: Popular mixed matchings. Theoretical Computer Science 412, 2679-2690 (2011)

24. Brandt, F., Hofbauer, J., Suderland, M.: Majority graphs of assignment problems and properties of popular random assignments. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, pp. 335-343. (2017)
25. Abdulkadiroğlu, A., Pathak, P., Roth, A.: Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. The American Economic Review 88, 1954-1978 (2009)
26. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29-36 (1982)
27. Diebold, F., Bichler, M.: Matching with indifferences: A comparison of algorithms in the context of course allocation. European Journal of Operational Research 260, 268-282 (2017)
28. Aziz, H., Brandt, F., Brill, M.: The computational complexity of random serial dictatorship. Economics Letters 121, 341-345 (2013)
29. Che, Y.-K., Kojima, F.: Asymptotic equivalence of probabilistic serial and random priority mechanisms. Econometrica 78, 1625-1672 (2010)
30. Katta, A.-K., Sethuraman, J.: A solution to the random assignment problem on the full preference domain. Journal of Economic theory 131, 231-250 (2006)

Applied image recognition: guidelines for using deep learning models in practice

Matthias Griebel[1], Alexander Dürr[1], and Nikolai Stein[1]

[1] University of Würzburg, Department of Business and Economics, Würzburg, Germany
{matthias.griebel, alexander.duerr, nikolai.stein}
@uni-wuerzburg.de

Abstract. In recent years, novel deep learning techniques, greater data availability, and a significant growth in computing powers have enabled AI researchers to tackle problems that had remained unassailable for many years. Furthermore, the advent of comprehensive AI frameworks offers the unique opportunity for adopting these new tools in applied fields. Information systems research can play a vital role in bridging the gap to practice. To this end, we conceptualize guidelines for applied image recognition spanning task definition, neural net configuration and training procedures. We showcase our guidelines by means of a biomedical research project for image recognition.

Keywords: Deep learning, Image Recognition, Object Detection, Instance Segmentation, Artificial Intelligence.

1 Introduction

In recent years, novel deep learning techniques, greater data availability, and a significant growth in computing powers have enabled AI researchers to tackle problems that had remained unassailable for many years. This holds especially true for voice or image recognition tasks where deep learning has demonstrated its remarkable capability of revealing structures in unstructured high-dimensional data. Given the wide availability of such data, deep learning applications can be used in many areas of science, business and administration [1]. At this point, a *McKinsey* study estimates the potential of AI applications to create between $3.5 trillion and $5.8 trillion in value annually across nine business functions in 19 industries [2].

A case in point for image recognition applications is the health care sector where deep learning in conjunction may offer a critical complement to the gold standard of randomized controlled trials by supporting massive observational studies that were not feasible before [3]. While there are already many successful biomedical applications enabled by deep learning applications, there is still a great need for innovative solutions. *Grand Challenge* [4] lists 167 data science competitions for biomedical image analysis over the last decade. These challenges comprise a wide range of applications, from ultra sound nerve segmentation, determination of skeletal age, and multiple sclerosis segmentation to different sorts of cancer detection and classification. A recent example

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

is the *Kaggle Data Science Bowl 2018* [5] that aims to develop algorithms to speed up research for almost every disease, from lung cancer and heart disease to rare disorders or to the common cold. While the IS community actively engages in various healthcare-oriented fields such as health care management [6], health care services [7] or mental health therapy programs [8], there has been little activity towards supporting researchers with cutting-edge tools such as advanced image recognition. Yet, our community should assume a more active role in this field as it is "uniquely positioned to provide the appropriate mix of rigor along with humanistic and instrumental relevance" [9].

In recent years, comprehensive new AI frameworks such as *Keras* [10] have emerged. They focus on fast experimentation and prototyping through user-friendliness, modularity, and extensibility. The corresponding democratization of AI allows non-AI researches to easily access powerful deep learning applications. This shifts the focus of attention from the technology to the use case. We feel that this development offers a unique opportunity for information systems researchers in facilitating the use of these tools in practical applications. Alongside this development, the availability of unstructured data, notably image data, is increasing dramatically. Images are not only present on social media platforms (Instagram, Facebook), video platforms (YouTube), satellite images (such as Planet.com), but also a growing constituent in scientific research [11]. As the volume of image data has vastly exceeded the capacity of manual analysis, AI is henceforth a key component for automated evaluation [12].

For research purposes AI applications, as with traditional machine learning applications, are typically embedded in data mining pipelines. Existing data mining frameworks such as the guidelines put forward by Müller et al. [13] or CRISP-DM [14] only vaguely describe machine learning applications as part of the *modeling* phase, whereas they focus on tasks such as feature engineering in the data *preparation* phase and the data mining process itself. However, modeling is a critical and extremely complex task for the distinctive nature of deep learning (AI) methods.

To this end, we seek to outline the current state of advanced image recognition and contribute to the literature by providing tangible guidelines for non-AI researchers on how to incorporate state-of-the-art AI algorithms into data mining pipelines. Thereby, we follow up on the call for embracing the value of unstructured data in the design of analytical information system put forward by Müller et al. [13].

2 Building Blocks for Image Recognition Applications

Supervised learning for image recognition requires a data set of labeled images (e.g., magnetic resonance or microscopy images labelled healthy or infected) [1].[1] To facilitate the usage by researchers outside the AI world we want to establish general guidelines for setting up computer vision projects.

[1] In the following, we assume the availability of such data and do not address the also challenging collection task in the remainder of this paper.

To this end, we break down the image recognition into its main building blocks - task definition, the design of the neural network and finally the training approach. The design task features several sub-tasks (choice of architecture, loss function, evaluation metric). To offer concise recommendations for these highly technical sub-tasks we link the design of the neural net to the initial task definition.

2.1 Defining the Task

In order to effectively address the abundance of image recognition applications, it is imperative to understand the underlying problem set. Consequently, any applied computer vision project must ultimately start with a proper definition of the image recognition task at hand. The majority of applications are capture by the following main task categories:

- *Image classification* [15] assigns the whole image to a particular class.
- *Semantic segmentation* (also referred to as pixel-wise classification) identifies every pixel that is part of a specific class, while neglecting distinct instances [12], [17].
- *Object classification* (also referred to as object detection) distinguishes between different objects (instances) of classes in a picture, returning their approximate location using a bounding box [18].
- *Instance segmentation* localizes objects on a pixel basis [19].

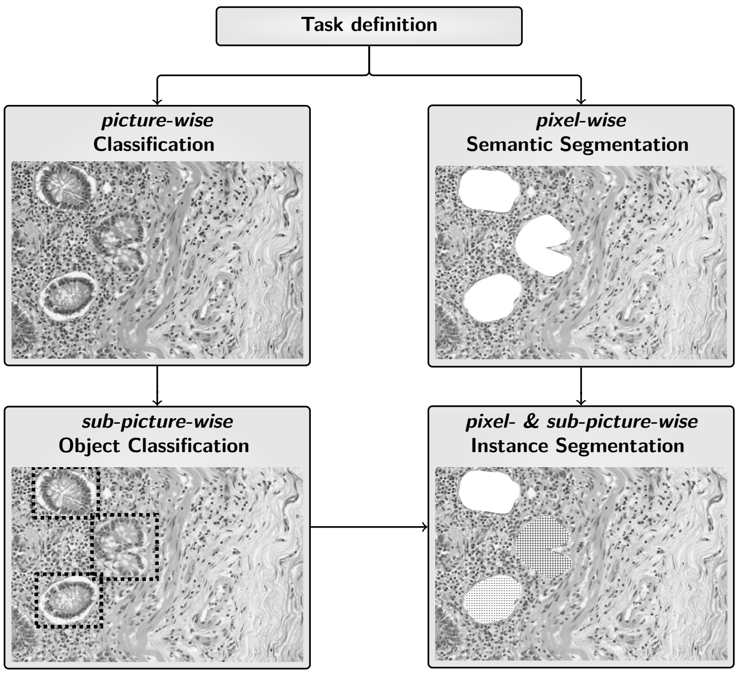


Figure 1. Taxonomy of image recognition tasks using the example of histology images that show cancer cells.

We want to illustrate these categories by means of a histology image containing cancer cells (Fig. 1). The histology images were adapted from "The GlaS Challenge Contest" data set [20]. Depending on the focus of the study the following questions can be addressed using image recognition:

- Classification: Does this image contain any cancer cells? If yes, assign this image to the class “cancer”.
- Semantic Segmentation: What pixels belong to the class “cancer”?
- Object Classification: How many cancer cells are in the image and what is their approximate location?
- Instance segmentation: How many cancer cells are in the image and what is the exact (pixel) position?

2.2 Composing the Neural Network

Having identified the image recognition task, the underlying neural network for image analysis must be set up. Unlike other classification or regression techniques, this is a highly non-trivial task and requires interacting with oftentimes cryptic concepts and an overwhelming number of design options.

While artificial neural networks, i.e., multilayer perceptrons, have been successfully applied to various tasks since the 1980s, convolutional neural networks (CNN) have emerged as the standard for image recognition in the last decade [1]. Consequently, we focus on explaining the essential building blocks of this class of neural networks and establish best practices for each task category.

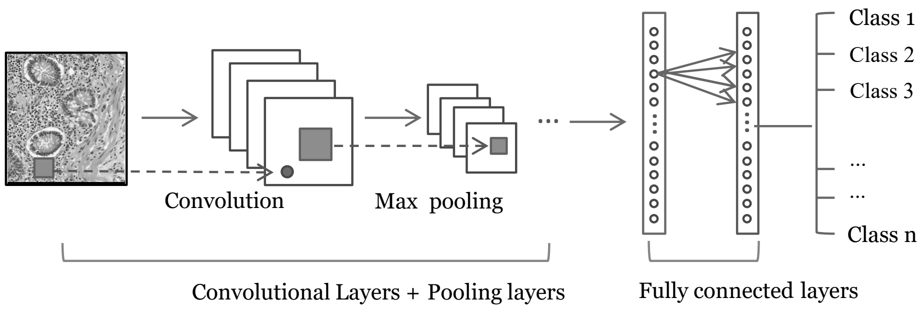


Figure 2. Example of a CNN Architecture

Architectures for Convolutional Neural Networks. The general CNN architecture is composed of three main neural layers, namely convolutional, pooling and fully connected layers as shown in Fig. 2 [1]. Convolutional layers consist of filters (“neurons”) and feature maps to discover conspicuous local pattern-like edges, lines, and other visual elements. Pooling layers are typically considered as a technique to compress or generalize feature representations and reduce the overfitting on the training data by the model [21]. Fully connected layers are used at the end of the network after feature extraction and consolidation by the convolutional and pooling layers. They integrate all feature responses and provide the final classification results [21].

The overwhelming success of AlexNet [21], a large CNN for image classification, in the ILSVRC 2012 challenge [22] has sparked significant interest in the CNN approach. Since then a vast number of architecture tweaks have emerged, each offering incremental improvements of image classification for different data sets. By using their original configuration, these networks perform the task of image classification. Due to their remarkable ability to extract features from images, they are also used as a *backbone* architecture for other image recognition tasks. Fig. 3 provides an overview of the current main architecture choices.

The basic VGG family, introduced by [23], is typically used for its simple and easily understandable architecture (see Fig. 3).

The *Inception* family of networks (introduced by [24]) relies on Inception modules (Fig. 3), where the input is processed by several parallel convolutional layers of different sizes whose outputs are then merged back. This enables the network itself to converge towards an optimal level of abstraction to represent a feature. Finally, the *ResNet* family [25] introduces residual blocks to the CNN (see Fig. 3). Their special features are shortcut connections parallel to the convolutional layers. This facilitates the efficient training of even deeper and more powerful networks [25]. Moreover, these architectures are frequently used as a foundation to tailor *customized* CNN architectures towards a specific use.

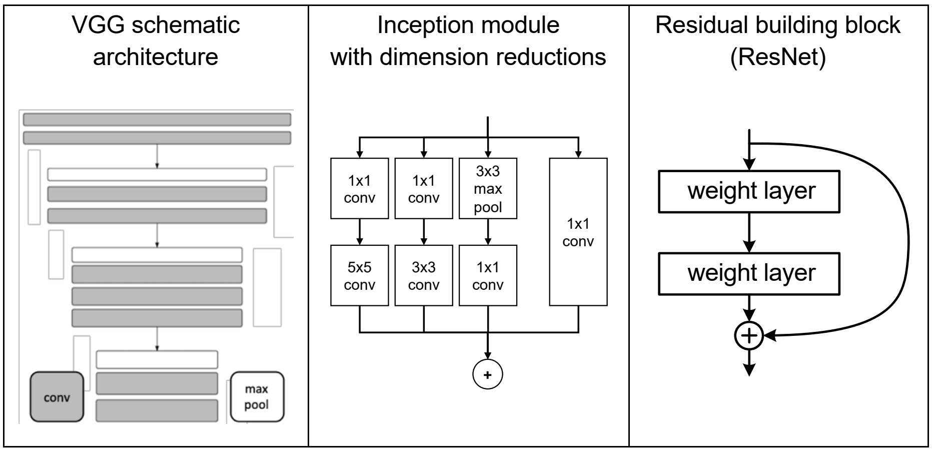


Figure 3. Backbone architecture characteristics

Next, we want to match CNN architectures to the image recognition task categories. These suggestions should provide an informative starting point for determining a suitable architecture:

- Classification: At present, the best performing classification models are, e.g., Inception-Resnet-V2 [26] different version of ResNet (i.e., ResNet51, ResNet101) or VGG.

- Semantic segmentation: Depending on the purpose, variants of the U-Net [27] (using a VGG backbone) perform well on biomedical images such as 2D light microscopy cell segmentation. The 3D version of the U-Net is called V-Net [28]. For more general purpose applications we suggest a VGG based architecture such as a Fully Convolutional Network (FCN) [16].
- Object classification and instance segmentation: As the approach of Mask R-CNN [19] allows both object detection and instance segmentation within the same setting it is the best option for most multi-class segmentation applications. However, the U-Net variants can be extended by additional post-processing steps to enable instance segmentation. In particular, this approach showed very strong performance in the *Kaggle Data Science Bowl 2018* [5]. Depending on the problem at hand, it can be rewarding to implement and evaluate both approaches.

Loss Function and Optimizer. The loss function (objective) and optimizer are the main components to configuring the learning process of a neural network. During the learning phase the weights are adjusted so that the loss decreases. The loss function has to be chosen according to the task, the number of classes or potential class imbalances. Due to its robustness and ability to handle nonlinear effects, the binary cross entropy loss is commonly used as the standard loss for binary classification tasks (picture- or sub-picture-wise). Accordingly, the categorical cross entropy loss works well for all multi-class classification tasks.

In pixel-wise segmentation tasks there is typically an imbalance between pixel classes (i.e., many background pixels and few foreground pixels). There are two common approaches to cope with this problem. On the one hand, [27] propose the use of the weighted cross entropy loss. On the other hand, the dice coefficient loss yields promising results as it handles true negatives as uninteresting defaults [28].

The optimizer determines the update process of the CNN by calculating the gradient. To tackle the high volumes of image recognition tasks, it is of paramount importance that the optimizer computational efficient, has little memory usage and requires little tuning. We suggest to use the optimizer Adam as it outperforms other common choices (e.g., SGD, AdaDelta and RMSProp) with respect to computational overhead [29].

Evaluation Metrics. A suitable evaluation metric is needed to assess a model's performance on the image recognition task. In contrast to the loss function, metrics do neither require to be mathematically differentiable nor used to train the model. Understanding the importance of the evaluation metric is fundamental for every data science project [30], including image recognition tasks.

Accuracy and the area under the curve (AUC) are metrics to evaluate the quality of classification results. For class-imbalanced problems, the Mathew correlation coefficient (MCC) is considered a robust measure [31]. Recall, precision and F-Measure focus on the positive examples to capture information about the rates and kinds of errors made. The intersection-over-union (IoU) metric measures the similarity between the predicted region and the ground-truth region for an object present in the set of images. This is particularly suited for pixel-wise image segmentation tasks. There is clearly no gold standard for evaluation metrics, as they have to account for the specific properties of the given task and underlying data set. We suggest using a combination of different metrics in order to cover different aspects of the evaluation requirements. An exemplary combination of metrics for instance segmentation could be the IoU and recall. While the IoU measures the quality of the segmentation task, the recall accounts for the ability to detect all relevant instances.

2.3 Training Strategy

Having determined the composition of the neural network (by choosing an appropriate CNN architecture, loss function, optimizer and evaluation metric) the final task of training this network on the data needs to be tackled. To this end, we introduce different concepts and best practices for model generalization, hyperparameter optimization, and hardware requirements.

Model Generalization. The advantage of deep and complex CNN architectures is to better extract information from unstructured data. However, a large number of available parameters (weights) renders these networks prone to overfitting which prevents the model from generalizing well to unseen instances [32]. We consider *data-oriented techniques*, *transfer learning*, and *architectural tweaks* to limit the overfitting tendencies of a model.

Data-oriented techniques prevent overfitting by restricting full access of the network to the training data. To this end we apply methods such as data splitting and data augmentation. Data splitting partitions the data set into two subsets: training and validation. The model is then trained on the training data and evaluated on the validation data. Thus, it is possible to stop the training as soon as overfitting occurs. In a k-fold cross validation this procedure is repeated k times [33].

Data augmentation artificially generates additional data without incurring extra labeling costs. In the case of image recognition this is easily achieved by means of transformative methods, such as rotation, shearing, translation, flipping, elastic deformations, and random intensity jitter. This is especially useful for small data sets [21]. Depending on the data set, some transformations should not be performed, i.e., in case of an object recognition task where objects are characterized by their shape, the shape should not be distorted.

Moreover, transfer learning leverages a pre-trained model as feature extractor. To this end, the CNN is initialized with pre-trained parameters of a network that has been trained on another data set such as ImageNet [22] or MS COCO [34]. There are plenty of pre-trained models publicly available, e.g., in the repository of *Keras* [10]. The pre-trained model is fine-tuned subsequently. Thereby, the pre-trained parameters of the initialized network are gradually adjusted to the new images during additional training steps. Depending on the problem, oftentimes the parameters of the majority of the layers are fixed while only a few parameters on top layers[2] are adjusted. Optionally, some custom layers can be introduced and trained in parallel to fine-tune these layers

2 Here, we define *top layers* as the layers close to the input layer of the CNN.

on the new data set. In general, transfer learning accelerates the learning process and improves the generalization ability of a network [32].

Finally, architectural considerations such as dropout layers [35] can incorporate generalization approaches within the CNN composition. Dropout layers prevent the network from overfitting by randomly deactivating a share of the neurons during the training phase. Thereby, the model is forced to learn the same patterns using different neurons. During the prediction phase the dropout is deactivated and all neurons can be utilized.

Hyperparameter Optimization. There are numerous configuration settings in a CNN that can be tuned to improve the performance. Such parameters include, e.g., the activation function, learning rate, the number of training epochs, the batch size, the, initial weight choices and many more.

- Each weight layer in a CNN is typically ensued by a non-linear *activation function*. The simplest activation function for binary classification decisions is the sigmoid function which is bounded between 0 and 1. The ReLU (Rectified Linear Unit) activation function [36] is commonly used for all layers except for the output layer in practice because of the constant slope for positive values.
- The *learning rate* controls the magnitude weights adjustment after each iteration. If the learning rate is low, the training progresses slowly. In contrast, a high learning rate can prevent from converging to a possible minimum loss.
- One *epoch* is when an entire dataset is passed through the neural network for training.
- The *batch size* defines the number of samples propagated through the network in each step of gradient descent, i.e., learning.

Given the vast number of parameters manual tuning is impossible. Consequently, we suggest to conduct an automated hyperparameter search based on either a random grid search [37] or a Bayesian optimization search [38] to identify the promising parameter choices.

Hardware Requirements. The training of CNNs requires a vast number of convolutional operations resulting in an enormous demand for computing power. Training the model on purpose-built hardware such as GPUs or TPUs are far more efficient than training on a universal CPU. The increased availability and reliability of cloud-computing services provides a strategic dynamic capability to scale up or down the IT infrastructure [39]. Therefore, we suggest using Machine Learning as a Service (MLaaS) solutions. Such services are offered by all leading cloud operators.

3 Applying the Guidelines: A Biomedical Case Study

We illustrate the execution of an image recognition project based on the guidelines put forward above. To this end, we report learnings from a research collaboration with a group of neuroscientists. In a joint project we developed an data mining pipeline to automatically detect fluorescently stained neurons in tissue images of mice brains [40].

3.1 Defining the Task

To define the task, we first need to understand the underlying problem and data set. Fig. 4 shows an excerpt of image dataset obtained using a confocal microscope. The data comprises three different sub-regions of the dorsal hippocampus: dentate gyrus (DG), Cornu ammonis 1 (CA1) and CA3. As there is no ground truth for fluorescent signal segmentation, neurons are determined by their relative brightness (signal strength) to the background. For this purpose, the resulting segmentation maps are generated either by means of a heuristic, manual identification process or by means of a (partially) automated threshold-based analysis. Due to the low signal-to-noise ratio of the data, threshold-based approaches do not work reliably as they fail to detect most of the fluorescent areas (see Fig. 4).

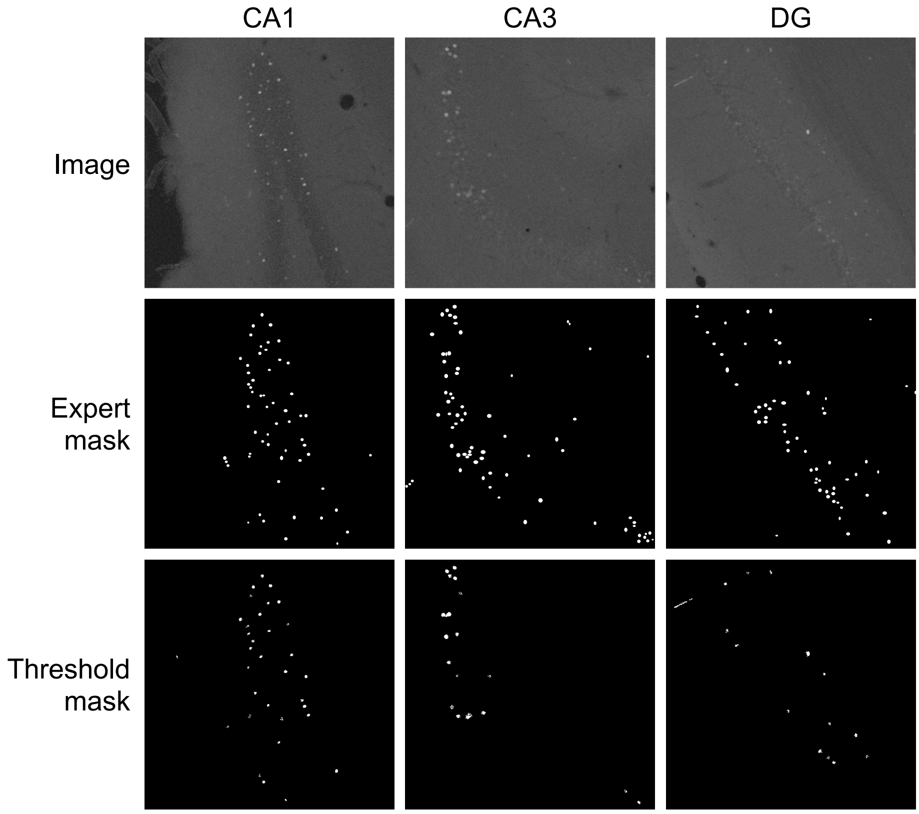


Figure 4. Different sub-regions of the dorsal hippocampus and the corresponding segmentation masks (here, the threshold only considers the 5% brightest pixels per image).

The goal of our image recognition is to automatically detect fluorescent neurons within a microscopy image. For biomedical evaluation, researches require the position, size and signal intensity of fluorescent neurons. Thus, our model needs to identify (i) object instances as well as (ii) the exact area (segmentation mask) rendering instance segmentation suitable for our task.

3.2 Composing the Neural Network

CNN Architecture. According to the task definition we first used the Mask R-CNN approach based on a ResNet backbone architecture for instance segmentation. This already yielded reasonably good results but also required a huge amount of computational resources. We also tried a U-Net based approach similar to the winning solution of the *Kaggle Data Science Bowl 2018* [5]. In this particular case, instance segmentation is achieved by (i) performing pixel-wise binary classification with the U-Net and (ii) post-processing the resulting binary segmentation map. The post-processing pipeline includes a Watershed algorithm [41] and the removal of biological implausible regions (i.e., too small or misshapen). As the U-Net approach yields the better results we continue with it for the remainder of the study.

Loss and Optimizer. As shown in Fig. 4 the total number of fluorescent neurons (positive pixel class) is far less than the background (negative pixel class) resulting in high class imbalances among the whole dataset. Thus, we optimize (Adam algorithm) our model by minimizing a weighted combination of the cross-entropy loss and the dice loss to take advantage of their respective benefits. Here, the dice coefficient loss is particularly valuable as it handles true negatives as uninteresting defaults. We found that the outcome of the whole pipeline depends on a well-suited loss function.

Metrics. To evaluate the quality of our model we compare the expert segmentation masks to the post-processed output masks of our network. This comparison can either be performed pixel-wise or on an aggregated neuron level. For the pixel-wise comparison we need to take the class imbalance into account. Hence, we leverage the IoU as we are mainly interested in identifying instances of the positive class (fluorescent neurons).

Considering the biomedical use case, researches are particularly interested in position and size of each neuron. However, in high resolution images the exact boundaries of the neurons are difficult to define for human experts on a pixel level. As a result, there are often minimal deviations on pixel level even though the same neuron is detected. To address this issue, we introduce another comparison process that (i) matches the corresponding neurons of two segmentation masks and (ii) calculates the accuracy as the proportion of matches divided by the total number of unique neurons found on both segmentation masks.

3.3 Training Strategy

Due to the high cost for both manual labeling and mice experiments, only a limited amount of training samples are available. Thus, we apply data augmentation as a combination of randomly rotating, flipping and shifting the original image-mask pairs. As the shape of the neurons is important in the identification process, we do not use techniques that distort the shape (e.g., shearing). Fig. 5 exemplifies this process with random parameters. Here, the original image-mask pair is horizontally flipped, rotated by 90 degrees clockwise and 20 percent shifted to the top and right. In light of the small dataset, data augmentation prevents from overfitting and generalizes the model, e.g. by learning to detect neurons independent of their position.

To further remedy the issue of limited training data we pre-trained our model on the Kaggle Data Science Bowl 2018 [5] data set, which contains similar microscopy tissue images. To tune the parameters of the network we use a Bayesian optimization search. The model is trained and evaluated on multiple Nvidia Tesla V-100 GPUs.

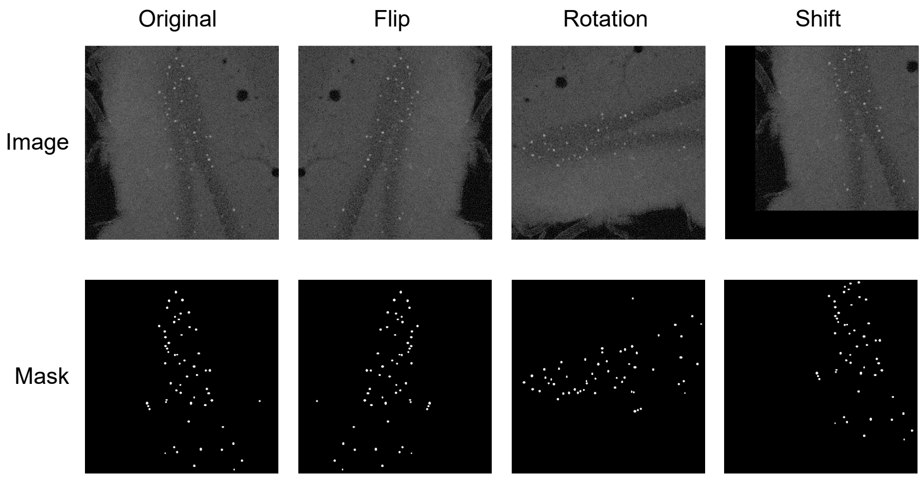


Figure 5. Data augmentation methods used in our project.

The detailed findings of our research project are described by Segebarth et al. [40] and the code is publicly available on *GitHub*[3]. In order to communicate our research, we provide a Jupyter-Notebook with free access to high computing power on *Google Colab*[4]. The execution of the notebook requires no machine learning and almost no programming expertise.

[3] https://github.com/matjesg/DeepFlaSH

[4] https://colab.research.google.com/

4 Conclusion and Outlook

In this study, we outline the current state of advanced image recognition and provide guidelines for non-AI researchers on how to incorporate state-of-the-art AI algorithms into data mining pipelines. We showcase the application of the proposed guidelines on a case study in the field of biomedical image recognition.

In particular, our research aims to make several contributions to the literature. First, we structure the variety of existing image recognition approaches and put forward a taxonomy for image recognition tasks based on the relevant literature. Second, our proposed guidelines are expected to extend data mining frameworks such as CRISP-DM. The *modeling* phase for machine learning applications is only vaguely described here, although it is a complex challenge for deep learning models. Finally, our presented case study demonstrates the potential of AI in biomedical applications. The automation of the conventional (manual) image analysis process reduces workload and enables highly qualified researchers to focus on important activities instead of tedious image labeling work.

To tap into these benefits, some limitations of our guidelines need to be considered. In the rapidly materializing field of AI and deep learning, the proposed model architectures and recommendations only represent a snapshot of the current developments and need to be updated continuously. However, the taxonomy for image recognition tasks will also apply to new technologies. In addition, modeling deep learning applications may require additional specific domain knowledge as depicted in our case study. Another difficulty is to identify the appropriate degree of abstraction for our guidelines when addressing scholars with different levels of prior knowledge.

In future research, we plan to generalize and refine our proposed guidelines by means of an evaluation on use cases from different domains. Possible topics may comprise but are not limited to fashion trend detection, satellite image analysis to predict future resource requirements and the diagnosis and management of diseases.

References

1. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature. 521, 436–444 (2015)
2. Parker, K., Stone, J.A., Arena, R., Lundberg, D., Aggarwal, S., Goodhart, D., Traboulsi, M.: Notes from the AI frontier: Insights from use cases. McKinsey Global Institute (2018)
3. Agarwal, R., Dhar, V.: Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Inf. Syst. Res. 25, 443–448 (2014)
4. Grand Challenges, https://grand-challenge.org/All_Challenges/ (Accessed: 10.10.2018)
5. Kaggle Data Science Bowl, https://www.kaggle.com/c/data-science-bowl-2018/ (Accessed: 10.10.2018)
6. Wager, K.A., Lee, F.W., Glaser, J.P.: Health care information systems: a practical approach for health care management. John Wiley & Sons (2017)
7. Yaraghi, N., Du, A.Y., Sharman, R., Gopal, R.D., Ramesh, R.: Health Information Exchange as a Multisided Platform: Adoption, Usage, and Practice Involvement in Service Co-Production. Inf. Syst. Res. 26, 1–18 (2015)

8. Lederman, R., Wadley, G., Gleeson, J., Bendall, S., Álvarez-Jiménez, M.: Moderated online social therapy: Designing and evaluating technology for mental health. ACM Trans. Comput.-Hum. Interact. 21, 1–26 (2014)
9. Abbasi, A., Sarker, S., Chiang, R.H.L.: Big data research in information systems: Toward an inclusive research agenda. J. Assoc. Inf. Syst. 17, 1–32 (2016)
10. Chollet, F., others: Keras, https://keras.io (Accessed: 10.10.2018)
11. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. 36, 1165–1188 (2012)
12. Provost, F., Fawcett, T.: Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. 1, 51–59 (2013)
13. Müller, O., Junglas, I., Brocke, J.V., Debortoli, S.: Utilizing big data analytics for information systems research: Challenges, promises and guidelines. Eur. J. Inf. Syst. 25, 289–302 (2016)
14. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0 Step-by-step data mining guide. (2000)
15. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621 (1973)
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
17. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528 (2015)
18. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
19. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
20. Sirinukunwattana, K., Pluim, J.P.W., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., Böhm, A., Ronneberger, O., Cheikh, B.B., Racoceanu, D., Kainz, P., Pfeiffer, M., Urschler, M., Snead, D.R.J., Rajpoot, N.M.: Gland segmentation in colon histology images: The glas challenge contest. Med. Image Anal. 35, 489–502 (2017)
21. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)
22. Russakovsky, O., Deng, J., Krause, J., Berg, A., Fei-Fei, L.: The ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. IJCV. 115, 211–252 (2012)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. ArXiv (2014)
24. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
25. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE, Las Vegas, NV, USA (2016)
26. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Pattern Recognit. Lett. 42, 11–24 (2016)

27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241 (2015)
28. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th International Conference on 3D Vision, pp. 565–571 (2016)
29. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. ArXiv 1–15 (2014)
30. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML 2007, pp. 209–216 (2007)
31. Powers, D.M.: Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. J. Mach. Learn. Technol. 2, 37–63 (2011)
32. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: A review. Neurocomputing. 187, 27–48 (2016)
33. Kohavi, R.: A study of Cross validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the International Joint Conference on Neural Networks, pp. 1137–1145 (1995)
34. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3686–3693 (2014)
35. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J. Mach. Learn. Res. 15, 1929–1958 (2014)
36. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814 (2010)
37. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305 (2012)
38. Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., Sculley, D.: Google Vizier. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1487–1495 (2017)
39. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.: Digital business strategy: toward a next generation of insights. MIS Q. 37, (2013)
40. Segebarth, D., Griebel, M., Duerr, A., R. von Collenberg, C., Martin, C., Fiedler, D., Comeras, L.B., Sah, A., Stein, N., Gupta, R., Sasi, M., Lange, M.D., Tasan, R.O., Singewald, N., Pape, H.-C., Sendtner, M., Flath, C.M., Blum, R.: DeepFLaSh, a deep learning pipeline for segmentation of fluorescent labels in microscopy images. bioRxiv (2018)
41. Beucher, S.: Use of watersheds in contour detection. In: Proceedings of the International Workshop on Image Processing. CCETT (1979)

Yield Prognosis for the Agrarian Management of Vineyards using Deep Learning for Object Counting

Kai Heinrich[1], Andreas Roth[1], Lukas Breithaupt[1], Björn Möller[1], Johannes Maresch[1]

[1] Technische Universität Dresden
{kai.heinrich,andreas.roth,lukas.breithaupt,
bjoern.moeller,johannes.maresch}@tu-dresden.de

Abstract. In various applications, the counting of objects based on image data plays a pivotal role. In this paper we first conducted a literature review to display the state of the art in counting objects and summarized the results by extracting several important concepts that describe the counting problem as well as the solution. In a second step we applied this knowledge to yield prognosis in vineyards, where we used Deep Learning models to detect the objects. While these methods used in the detection step are state of the art and perform very well, several problems are usually introduced by the constraint of only counting an object once in the counting step. We provide a solution for this common problem by identifying unique objects and tracking them throughout a sequence of images in order to avoid counting objects more than once, resulting in an automated yield prognosis model for vineyards.

Keywords: deep learning, counting, image data, vineyard management

1 Introduction

The forecasting of economical features like profit plays an important role in predictive analytics in the age of digitization. In addition to the service and manufacturing industry, the commodities sectors, especially the field of agrarian management can benefit a lot from disruptive and innovative processes that follow up the digital transformation process [1]. Especially the scientific field of artificial (AI) with its subsidiary field of machine learning (ML) aims to provide a solution to problems that are either very cumbersome for humans to tackle or simply impossible [2]. However, there are still some fields, where artificial intelligence can help to tackle problems that are otherwise cumbersome to tackle.

One of these fields is yield prognosis in the context of agrarian management [3, 4]. The yield prognosis task within vineyard management is often times very labor intensive and subject to errors generated by bad and subjective sampling [5]. In addition to that the manual labor sampling techniques have very poor scaling behavior towards larger vineyards. While there are some automatic approaches using complex lighting

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

camera techniques to provide image data for statistical models, they are for the most part even more costly than the actual manual labor sampling process and have to be reapplied every time since the models are subject to various spatial and temporal factors like terrain, season and time of day.

While the problem of decision support systems that excel in the task of predicting profits based on yield prognosis is well established in IS-research, companies often times fail to leverage new analytic tools to harness business values from growing amounts of data [6–8]. While most systems promise improved results in terms of prediction performance, they come with high barriers in terms of application interpretability and required computational power [9, 10]. Therefore, the construction and evaluation of AI technologies like deep learning systems (DLS) in the context of business problems should be considered a vital research interest in IS research [11]. We construct such systems by transferring known algorithms for object counting to the problem of yield prognosis.

The task at hand can therefore be reformulated as finding a model that is able to count agrarian entities like grape berries and whole plants of grape vines. This task adds an additional and important step in the data analysis process: Besides the usual description of data selection, data preparation, modeling and evaluation we add a counting step after the model phase since basic detection is not enough to provide a yield prognosis.

Our goal is to provide several solutions to solve the underlying problems of object detection (modeling phase) and yield prognosis (counting phase) and we therefore employ a two-step-approach: Following a proposed IS research gap by Gordon et al. 2013, we review related work on intelligent systems based on Artificial Neural Networks (ANN) for the task of both object detection and counting the identified objects, propose a systematization of the whole context and secondly evaluate state-of-the-art existing approaches for object recognition in combination with novel approaches for the counting step using image data collected from different vineyards throughout Europe. The paper is therefore structured as follows: In the next section we provide a description of the criteria for our literature review and the subsequent conceptualization of approaches for object counting.

We then proceed to describe the underlying data and analysis process for our data science study in Section 3. In Section 4 the results are provided and discussed with regard to the problem at hand. The last section provides a summary as well as research limitations and an outlook.

2 Preliminaries and related work

The task of counting objects based on image data has a wide variety of application domains, for example, counting cells on microscopic images or the surveillance of large people crowds as well as counting life stock or plants and trees from satellite or drone images [13]. Counting objects on various images is an exhausting and failure heavy process for humans, especially when the data sources contain multiple images (e.g., from video data) that needs to subject to automation for the reason of efficiency [14].

However using intelligent systems to count provides a lot of obstacles to overcome like hidden objects, fuzzy borders between objects or change of perspective [15]. Often times the counting process needs to provide additional information like the number of objects in a certain area (e.g., in air surveillance or agricultural analysis of plant positioning). Loy et al. differentiate between three basic approaches to counting: counting by detection, counting by segmentation and counting by regression [16]. Counting by detection is a two-step approach that first uses localization through object detection (e.g., through the use of a detection model like a convolutional neural network) and then proceeding to counting the detected objects. Counting by segmentation uses multiple images and groups consistent motions throughout the images to estimate the object number. Counting by regression is modeled after the human counting techniques for estimating or "guessing" object numbers without complete enumeration. The growing field of Artificial Intelligence (AI) and Deep learning (DL) enables the application of Artificial Neural networks (ANN), especially convolutional neural networks (CNN) for counting objects. While recent benchmarks suggest that counting by regression outperforms the standard counting by detection it is very cumbersome in the model development phase, since relevant features that help "guessing" the right number have to be crafted by hand for every situation. A CNN requires only labeled training sets and can learn higher order features by itself [17]. A well trained CNN can also be generalized in its architecture to be applied to different objects, e.g., a model that is used to count apples can be trained to count cars without changing the model structure [17]. Related surveys on these topics for both manually generated features and automated feature generation with deep neural networks are provided by [16] and [18] respectively. The surveys provide insights in the effectiveness of the different counting methods. It is suggested that standard counting by detection, while the most precise, fails to live up to the task in cases of images containing a large number of objects. The disadvantage of counting by segmentation is obviously that it cannot be applied to single picture data but needs a consistent stream of pictures with time stamps as they would be generated by videos. While counting by regressions seems to overcome both disadvantages, it often times requires the extraction of low-level features in order to estimate a density function on the domain of those features that reflects the density of objects for every pixel in an image. Integrating then leads to a cumulative function that produces an estimate of the total number of objects in an image [18]. However with the use of DL a certain degree of automation and general applicability can be achieved [19]. Before we evaluate some of the above mentioned approaches for the problem of yield prognosis in the agrarian management of vineyards, we provide a general overview over DL and ANN based algorithms for object counting by conducting a structured literature review.

3 Research Methodology and Literature Review

3.1 Research method

The research goal (RG) of this paper is (1) to provide an extensive state of the art overview of neural network based counting algorithms and (2) provide and evaluate a set of methods to overcome some of the problems that are inherent to counting by detection using data from a case of agrarian management of vineyards.

In order to provide a survey of related literature as formulated in RG (1) we employed a literature review with content analysis [20]. For the data science study formulated in RG (2), we use a study design as suggested by [21]. The structure of our data science study follows the KDDM approach by [22]. It is divided into six basic phases: *i) domain understanding*, *ii) data understanding*, iii) *data preparation*, iv) *modeling*, v) *evaluation*, and vi) *deployment*. Since the system we developed is a non-productive system without deployment we omit this phase from our research process.

As mentioned earlier we add a counting phase after the modeling phase, since using ANN algorithms to detect the objects is not enough. The counting step confronts us with the problem of unique entity counting, meaning that every object should only be counted once in order to provide a valid yield prognosis. We supply solutions to solve both problems – object detection as well as unique counting in our data science study for yield prognosis of vineyards.

For the identification of relevant literature we conducted a search with the terms "count" and either of the keywords "deep learning" or "neural network" in the databases Science Direct, EBSCOhost, IEEE Xplore and arXiv. Both the terms and the database choices cover the area of interest as suggested by RG (1). The search was conducted in a way, that the terms had to be present in at least one of the following meta data of the paper: abstract, title or keywords, where we explicitly defined the search string in a way that the term "count" had to be present in the title as a necessary condition. Initially 321 relevant papers were identified. We then proceeded to apply subjective filtering by screening the title (131) and then the abstract (87) for relevance. After removing duplicates and conducting a backward search we finally yield 99 relevant papers. A forward search based on the 99 papers revealed no additional results. The subjective filter was implemented based on some relevancy checks: (A) counting objects had to be a core topic of the paper, which prevented the inclusion object tracking or object recognition heavy papers; (B) the paper needed to be an original work rather than a survey, so that only papers that describe a method in-depth will be included in the overview; (C) the method was based on image rather than video data; (D) the article had to include neural networks (either standard or deep learning), since that was the focus of our research goal.

From our result set we were able to extract six core concepts that describe the various existing counting methods and their application environments: *architecture, number of objects, density of objects, background dynamics, output, training data* and *type of counting.* In the next section we provide the description of those concepts with the respective results from the literature review.

3.2 Literature Review results

Architecture. This concept distinguishes the counting methods based on extent and type of elements in the data analysis process responsible for counting. First we distinguish systems that only use one step processes (22/99), resulting in a single ANN that can be trained as a whole. Typically those networks were based on CNN architectures (e.g., [23–25]). Those methods typically use local and/or global features (Type of Counting is feature based) to determine the object density per pixel or partial and whole images. Another one step ANN involves classification whether an object is present in a given part of the image or not (Type of counting = Detection based) (e.g., [26]). Secondly there are counting systems that use multi steps that only involve ANN algorithms (16/99). Often times this extraction of partial images and then using it within a regression model (e.g., [27–29]). This machine based feature extraction provides advantages in scenarios with overlapping and strong variations in object size [30]. Other only ANN based multi-step approaches involve sequential models like Long-Term-Short-Term networks (e.g., [31]) or parallel architectures that use multiple networks as base learners while having a final "deciding network" in the second step (e.g. [30, 32, 33]). The last characteristic involves ANNs as well as other methods (61/99). This category acts as a collection for the versatile preprocessing possibilities of ANNs and other image processing methods. ANNs are used to either manipulate images to achieve better results (e.g., scaling or color related features), extract features (e.g. abstract or low level features like edges or texture) and for regression and combination of previous results.

Number of objects. This concept, while it should not be considered isolated, gives us a description of the counting problem at hand that we try to solve using ANN algorithms. The success of various counting methods depend largely on that concept, e.g., for counting of crowds of people we can use detection based methods for a small number of objects but not if we have large crowds present in images. We would then rather use feature based methods. We divided the found literature up into small number of objects ranging from 0-50 (55/99 papers), medium sized 50-200 (21/99 papers), large 200-1000 (14/99) and "outlier" or "extra-large" with the number of objects being larger than 1000 (4/99). Some methods were built to be more flexible and can be applied to various object numbers. In our review we took average numbers of objects based on the used image datasets. We could however not determine such numbers for five publications, since there was no indication regarding the number of objects.

Density of objects. The concept of density is another characteristic attribute of the counting problem and we can distinguish between no overlapping objects (31/99) and overlapping objects (68/99). It was found that for counting problems with considerable overlapping, feature based methods are preferred and more successful in predicting object numbers [13].

Background Dynamic. Another important influence factor found in the literature is the background dynamic which can be distinguished into static (43/99) and dynamic backgrounds (56/99). Static can have multiple meanings in this context: We can assume that over a sequence of images only the foreground changes (e.g., fixed surveillance cameras) where we can ignore the background and remove it via background subtraction methods (e.g., [34–36]) or as another option for static backgrounds we can assume a region of interest as a part of an image that is henceforth declared as foreground (e.g., [37, 38]). The assumption of static background is necessary for some methods to distinguish between foreground and background.

Output. The counting methods found in the literature offer a range of different outputs, depending on the specifics of the problem at hand. In some cases the methods only supply a rough estimate (10/99) without positioning or a total count. This is the case for public transportation surveillance where only a rough percentage estimate is needed, e.g., of how crowded a departure platform is (e.g., [39, 40]), reaching from 0-100%. Some methods output an estimate of the total number of objects in an image (59/99), again without positioning information. To circumvent this problem some approaches divide the image and count the total number of objects by region (8/99). With using overlapping regions, more robust models are generated (e.g., [14, 41]). Other output methods that take positioning into account involve determining a count per pixel by either returning a unique coordinate for each object centroid (2/99) or by using a bounding box that forms a rectangular shape around the object (4/99). Another approach to give an estimate per pixel is using a density function (15/99) as depicted by a heat map in Figure 1, where a higher pixel density is expressed by red color whereas low densities are expressed by blue color (e.g., [13, 25, 42]).

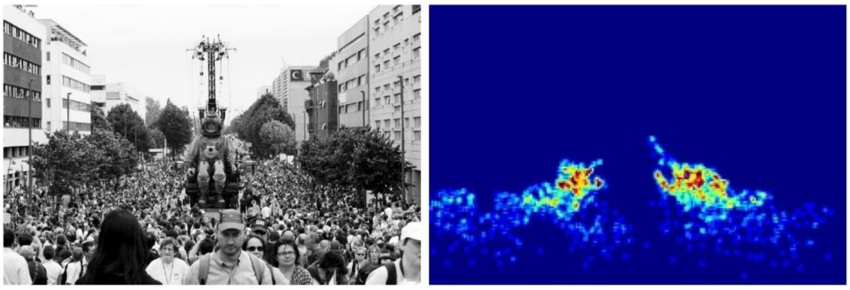


Figure 1. Example of density based count per pixel method ([28])

Training Data. The different options of available training data can be characterized by the effort of annotation of the unlabeled data. The first annotation tier with the lowest effort is given by annotating the total count (38/99). This is particularly useful when we only need an estimate or the total number as output without positioning information. For the purpose of annotating with positioning information we need the centroid coordinate (35/99), where every object is manually annotated with a dot in the center. This type of training data can be used for all output types since some outputs require the image to be divided up into regions, which in turn requires positioning information such as centroid coordinates (e.g., [28]). It was also found that this kind of data is needed in order to create density maps using Gaussian filter techniques (e.g. [30]). There are also approaches that do not require the annotation of training data at all (19/99) like segmentation based counting (e.g., [43, 44]) or detection based counting which instead of annotations needs example images of the objects that are to be counted (e.g., [44]).

Type of counting. The last concept as extracted from the literature review gives an insight on how the different approaches work. The first option is to use segmentation methods (15/99). After background and noise removal, which are essential preprocessing steps for this kind of approach, every foreground segment is declared as one object instance (e.g., [43]). The foreground/background segmentation is either conducted by a classification algorithm or by a threshold transforming images into binary black and white images with white pixels representing foreground objects [45]. The detection based methods (15/99) use a multi-step approach where objects are detected first and then localized and counted afterwards. Depending on the task at hand the detector is either trained to recognize the whole object (e.g. [44]) or only parts of it (e.g., only the head of a human body). The detectors are trained with single image training data to determine whether the object of interest is present or not. However a shortcoming of those approaches is that they cannot successfully be used with strong overlapping and large number of objects [16, 18]. Feature based methods (69/99) can circumvent that downside by extracting local and/or global features of the image. This can either be done by manually crafting features (e.g., [46]) or to use annotated data to train feature extractors (e.g., [23, 27]). Manually extracted features include shape, color, area, diameter, roughness of texture (e.g. [47]) or even additional meta annotations like perspective or weather data [48]. After feature extraction common machine learning algorithms like SVM, linear regression or ANNs are used to predict the desired output. In the next section we introduce a data science study that describes a data analysis process to provide an estimate for counting vines and grapes in order to produce a yield prediction to support decisions in vineyard management.

4 Data Science Study

4.1 Domain and Data Understanding

The management task in the context of agrarian management of vineyards is faced coherent spatial variability in the production systems, making crop and yield vigour an essential information [49]. Temporal variations of crop stability lead to unpredictable and nonstable yields, so that even the information of current yield during any phase of the crop development is vital.

In our data science study we gathered image data from three different vineyards in Germany at three different stages of the crop process (early, mid and late). The data gathering process was designed to be as cost efficient as possible, utilizing already existing equipment and avoiding complex lighting procedures and other high-cost image capturing in order to determine if a system can be designed that solves the counting task with only average to weak image quality data. We therefore combined already available mobile harvesting equipment like tractors with image capturing devices like full-HD cameras that generated a sequence of single images along a field of vine crops. We gathered three hours of video material in HD quality[1] and some Ultra-

[1] Taken with a GoPro Hero5 that was attached to the harvesting equipment

HD examples for testing purposes. Some important constraints when facing our task are a generalizable model that can predict yield under different conditions like weather, time of day and vineyard structure (e.g., slope or soil structure) and the necessity of avoiding to count an object more than once.

Since we can only use image data to train the model we transformed the video material into approximately 50 GB of image data sequences. We then judged the quality of the resulting images and especially the information content and removed poor quality images and images with no information content. Every image was captured from a frontal perspective facing the vines and we excluded other perspectives. We also excluded special vine crop structures like blue protection grids in order to keep generalizability (those vines have the same basic structure as other vine crops and are therefore detected just fine later on).

We can describe our counting problem in terms of the concepts extracted earlier in the literature review as a *small number of objects problem (0-50)* with *clear distinction* between the objects in terms of object density and for the most part *static backgrounds* with only small variations (e.g., cloud movement). We employ a *multi-step approach combining ANNs*. Since the object number is quiet high and we have some overlapping on the grapes, we use feature based detection. However we do not identify the image features ourselves but instead minimize the feature selection process by letting state-of-the-art image detection networks engineer the features, which further reduces manual labor on the part of the vineyard management. The concepts of output, training data and type of counting are given in the subsequent sections.

4.2 Preprocessing

For the labeling process we used the concept of *bounding boxes* for the training data as they are a supported format by deep learning frameworks like TensorFlow[2] and are our designated form of output for the modeling stages. We also labeled metal and wood rods that are occasionally used in vineyards in order to avoid confusing the algorithm. In other preprocessing steps we applied noise removal and feature extraction to highlight different parts of an image using thresholds and the background/foreground distinction resulting in black and white images, picturing vines with white pixels. For noise removal and isolation of objects of interest we further applied erosion on the transformed b/w images utilizing the OpenCV programming library and added Canny Edge Detection to better distinguish grape vines from the rest.[50]. Figure 2 gives a visualization of those preprocessing steps.

[2] https://www.tensorflow.org/

Figure 2. Selection criteria for input image data

For evaluation, we divided the data set into 70/30 train and validation data partitions.

4.3 Modeling

Given our counting task problem and the results of our literature review, we decided on a *detection based* type of counting by using automated feature extraction to minimize the manual effort in model building. There exist a lot of algorithms for efficient object detection and TensorFlow provides a collection of some of the state-of-the-art algorithms for that task [51]. We utilized the three common architectures for object detection ANNs: Single Shot Multi-Box Detector (SSD), Faster Region-Convolutional Neural Network (Faster R-CNN) and Region-based Fully Convolutional Networks (R-FCN). As a basis for our model training we used models that were pre-trained on the common benchmarking datasets: COCO (common objects in context), KITTI (images that capture traffic situations) and OID (an open Image dataset with 9 million images[3]).

We then continued training for the pre-trained models based on our training and test datasets. The results of the detection step in terms of model loss are given in Table 1. For evaluation purpose we compared the ground truth labeling bounding boxes with the model output boxes and used the mean average precision (mAP) as a metric [52]. In order for the comparison to be conclusive we use the degree of intersection between the two bounding boxes (ground truth vs. model output), also defined as Intersection over Union (IoU). The minimum threshold suggested by [52] is declared as an IoU of 50%. Since the original threshold was meant for more than 100 image object classes and we only have three classes (vine, metalstick, woodenstick), so we used an IoU of 70%.

Table 1. Object Detection Model Results

Model Architecture/ pre-training	Average Precision			Mean AP@0.5 IoU
	Vine	Wood	Metal	
F.R-CNN/coco	0.9887	0.9995	0.9992	0.9962
F.R.-CNN/kitti	0.8775	0.9081	0.9523	0.9126
F.R.-CNN/oid	0.9839	0.9991	0.9702	0.9847
SSD/coco	0.6027	0.6910	0.4433	0.5790
R-FCN/coco	0.9729	0.9982	0.9997	0.9910

[3] https://github.com/openimages/dataset

We can see that the task of vine detection executed successfully by all models except the SSD. For grape detection we therefore only used R-FCN and F.R-CNN models. Detecting single grapes was a nearly impossible task, since even with filtering the labeling of berries by humans was cumbersome and not always successful. We obtained average mAP values ranging from 0.2670 to 0.822, with the highest values only being reached when using some UHD quality images. Since obtaining and analyzing UHD images for a whole vineyard is beyond our economical reasonable goal definition of automated, cheap counting methods, we decided to cluster the berries up to grape vines and obtained mAP values ranging from 0.7576 to 0.9811 using an even higher IoU of 75%. The F.R-CNN/oid performed best with detecting grape vines.

4.4 Counting

As mentioned before we encounter several problems despite the successful detection of vines and grapes, since we cannot allow to count objects more than once. Using sequential image data we can define a simple predecessor – successor tracking process by giving an ID to a random box starting in frame 0 and then identifying that box in the next frame by drawing the same box from the previous frame and calculating overlapping percentage to all other boxes in the frame. The box with the most overlapping is called the successor and gets the same item ID. This creates a chain and identifies a single vine throughout the sequence. The process resets when there is no successor defined (when the end of the chain is reached) anymore and another box from frame 0 get a new ID. If all boxes already have an ID the process starts at frame 1 and so on, until all the vines have unique IDs. Sometimes the camera moves to fast or the camera reaches the end of the chain, so that we implemented a distance check that calculates the mean distance between two boxes. Whenever this distance is larger than the threshold of 150, no successor is defined, which lets the tracking process start again as explained above. The process is simplified and depicted in Figure 3 with the successor frames shown as red rectangles in frame n+1. A problem when using this method arises when we have (a) either missing vines in reality (e.g. vines that were removed because of crop sickness) or we (b) failed to detect a vine because it was hidden behind some other object (e.g. leaves or other plants).

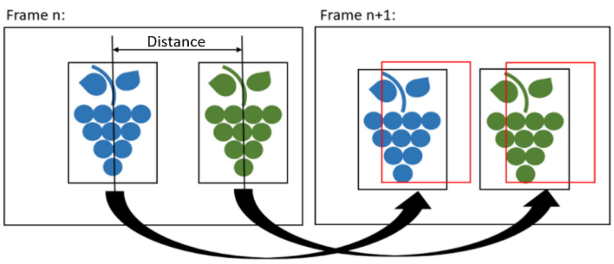


Figure 3. Tracking with overlapping and average distance between objects

In this case our method would not determine a successor object and counting would stop. But since we can track the vine throughout a sequence of images, we can determine the position the vine would have been at in that particular frame where case (a) or (b) applies. In case (a) we applied a 1.5 times threshold to the average distance measure as depicted in Figure 4 between two different IDs. For example if we calculate an average distance between two different vines (=different IDs) of 261 pixels, the threshold would be 391.5 pixels, resulting in the declaration of a missing vine whenever the distance between two bounding boxes exceeds that threshold. In case (b) we can correct the position where the successor should be by adding the average predecessor-successor difference to the previous position. If the object is out of bounce regarding to the frame coordinates we have simply reached the end of the row, otherwise we have imputed a vine which was not detected by our DL algorithm in the first place because of reason (b). This way we can solve the unique counting problem and also detect missing vines to give feedback to vineyard managers. This can have positive impacts especially when combined with agrarian management information like logs as to why vines were removed, so that we can prevent sickness from spreading or isolating certain crops in the development process.

We applied our process to every row we gathered video material from and the average deviation between model count and true vine count was 0.12 per row, with extreme values of 0 and 2. The average deviation for the count of grape vines per row was 1.87 with extreme values of 0 and 4. The extreme values can be explained by bad detection model performance on some objects, where the distance based algorithm just assumed there is either a missing vine or the end of row is reached. It is important to note that the counting process performance itself does not result in an error state itself, only when the object detection framework fails to classify an object. In comparison with other mechanics like motion tracking [25] or neural network based tracking [53] this combination is easier to use in the vine counting scenario as we described by using the extracted concepts from the literature review: medium number of objects, no overlapping, low density which is not the case in most benchmark scenarios of tracking in literature, where often times people are tracked in large crowds, facing problems of blur, high density and large amounts of overlapping, which requires far more advanced approaches [54]. However, in this case, it is sufficient in terms of the economic application to use a simple distance based method, since in comparison to the people tracking benchmarks we have a scenario with only slight perspective changes, distinguishable or static background and the overall scene is the same in every vineyard, apart from some details that have to be learned by the detection model.

5 Discussion & Outlook

We provided a literature review to describe both: the counting problem itself and the solution using various concepts like number of objects or type of counting. We then implemented a data science study to count objects like vines and grapes based on vineyard image data in order to provide a yield prognosis. While the detection of vines and grapes was very successful, the models failed to detect the single berries for non-

UHD picture material. We also provided a tracking solution for the counting step, to fulfill the constraint of unique object counting. However we only used a simple approach here that was based on subjective thresholds that might be subject to change on other vineyard architectures. Possibilities to circumvent this problem is the application of motion tracking straight onto the video material. While we made sure we had GPS data while gathering the image material we did not utilize it, which could be done in a next step and combined with aerial footage to map out the vineyard and provide more useful information like disease spreading factors to vineyard management, all of which would be automated and would not require human intervention once fully implemented. The process and the ideas provided in this paper are generally very robust to change of environment, especially the pre-trained networks can work on similar image data or can be re-trained in only a short amount of time, so that this process could be generalized onto various agrarian management situations were a similar counting problem structure can be found in terms of number of objects, density and background dynamics.

References

1. Shen, F., Chao, J., Zhao, J.: Forecasting Exchange Rate Using Deep Belief Networks and Conjugate Gradient Method. Neurocomput. 167, 243–253 (2015).
2. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press, Cambridge, Massachusetts (2016).
3. Font, D., Tresanchez, M., Martínez, D., Moreno, J., Clotet, E., Palacín, J.: Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. Sensors. 15, 8284–8301 (2015).
4. Fernandes, J.L.E., Nelson Francisco;Dalla Mora Esquerdo, Julio Cesar: Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. INTERNATIONAL JOURNAL OF REMOTE SENSING. 38, 4631–4644 (2017).
5. Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S.: Automated visual yield estimation in vineyards. Journal of Field Robotics. 31, 837–860 (2014).
6. Chen, H., Chiang, R.H.L., Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Q. 36, 1165–1188 (2012).
7. Sharma, R., Mithas, S., Kankanhalli, A.: Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. European Journal of Information Systems. 23, 433–441 (2014).
8. Shmueli, G., Koppius, O.R.: Predictive analytics in information systems research. Mis Quarterly. 553–572 (2011).
9. Arel, I., Rose, D.C., Karnowski, T.P.: Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE computational intelligence magazine. 5, 13–18 (2010).
10. Larochelle, H., Erhan, D., Courville, A., Bergstra, J., Bengio, Y.: An empirical evaluation of deep architectures on problems with many factors of variation. In: Proceedings of the 24th international conference on Machine learning. pp. 473–480. ACM (2007).
11. Agarwal, R., Dhar, V.: Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research. 25, 443–448 (2014).

12. Gordon, S., Blake, R., Shankaranarayanan, G.: Case-based research in information systems: Gaps and trends. JITTA: Journal of Information Technology Theory and Application. 14, 47 (2013).
13. Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: Advances in Neural Information Processing Systems. pp. 1324–1332 (2010).
14. Cohen, J.P., Boucher, G., Glastonbury, C.A., Lo, H.Z., Bengio, Y.: Count-ception: Counting by Fully Convolutional Redundant Counting. arXiv:1703.08710 [cs, stat]. (2017).
15. Zhang, C.: Design of Human Capability Maturity Analysis System Online P-CMM Model. Presented at the December (2015).
16. Loy, C.C., Chen, K., Gong, S., Xiang, T.: Crowd Counting and Profiling: Methodology and Evaluation. In: Ali, S., Nishino, K., Manocha, D., and Shah, M. (eds.) Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective. pp. 347–382. Springer New York, New York, NY (2013).
17. Seguí, S., Pujol, O., Vitrià, J.: Learning to count with deep object features. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 90–96 (2015).
18. Sindagi, V.A., Patel, V.M.: A survey of recent advances in CNN-based single image crowd counting and density estimation. Pattern Recognition Letters. (2017).
19. Goodfellow, I., Lee, H., Le, Q.V., Saxe, A., Ng, A.Y.: Measuring invariances in deep networks. In: Advances in neural information processing systems. pp. 646–654 (2009).
20. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a. MIS quarterly. 26, 13–23 (2002).
21. Müller, O., Junglas, I., vom Brocke, J., Debortoli, S.: Utilizing big data analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems. 25, 289–302 (2016).
22. Kurgan, L.A., Musilek, P.: A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review. 21, 1–24 (2006).
23. Capobianco, S., Marinai, S.: Deep neural networks for record counting in historical handwritten documents. Pattern Recognition Letters. (2017).
24. Hao Lu, Zhiguo Cao, Yang Xiao, Bohan Zhuang, Chunhua Shen: TasselNet: counting maize tassels in the wild via local counts regression network. Plant Methods. 13, 1–17 (2017).
25. Kang, D., Ma, Z., Chan, A.B.: Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks - Counting, Detection, and Tracking. arXiv:1705.10118 [cs]. (2017).
26. Achler, T., Vural, D.C., Amir, E.: Counting objects with biologically inspired regulatory-feedback networks. In: 2009 International Joint Conference on Neural Networks. pp. 36–40 (2009).
27. Aich, S., Stavness, I.: Leaf Counting with Deep Convolutional and Deconvolutional Networks. arXiv:1708.07570 [cs]. (2017).
28. Marsden, M., McGuinness, K., Little, S., Keogh, C.E., O'Connor, N.E.: People, Penguins and Petri Dishes: Adapting Object Counting Models To New Visual Domains And Object Types Without Forgetting. arXiv:1711.05586 [cs]. (2017).
29. Tang, S., Pan, Z., Zhou, X.: Low-Rank and Sparse Based Deep-Fusion Convolutional Neural Network for Crowd Counting. Mathematical Problems in Engineering. 1–11 (2017).
30. Boominathan, L., Kruthiventi, S.S.S., Babu, R.V.: CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. arXiv:1608.06197 [cs]. (2016).
31. Yao, H., Han, K., Wan, W., Hou, L.: Deep Spatial Regression Model for Image Crowd Counting. arXiv:1710.09757 [cs]. (2017).
32. Onoro-Rubio, D., López-Sastre, R.J.: Towards perspective-free object counting with deep learning. In: European Conference on Computer Vision. pp. 615–629. Springer (2016).

33. Zhang, Y., Zhou, D., Chen, S., Gao, S., Ma, Y.: Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 589–597 (2016).
34. Amin, I.J., Taylor, A.J., Junejo, F., Al-Habaibeh, A., Parkin, R.M.: Automated people-counting by using low-resolution infrared and visual cameras. Measurement. 41, 589–599 (2008).
35. French, G., Fisher, M., Mackiewicz, M., Needle, C.: Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video. Presented at the (2015).
36. Yoshinaga, S., Shimada, A., Taniguchi, R.: Real-time people counting using blob descriptor. Procedia - Social and Behavioral Sciences. 2, 143–152 (2010).
37. Laki, A., Rattalino, I., Corinto, F., Iván, K., Demarchi, D., Civera, P.: An integrated LOC hydrodynamic focuser with a CNN-based camera system for cell counting application. In: 2011 IEEE Biomedical Circuits and Systems Conference (BioCAS). pp. 301–304 (2011).
38. Hu, X., Zheng, H., Chen, Y., Chen, L.: Dense crowd counting based on perspective weight model using a fisheye camera. Optik - International Journal for Light and Electron Optics. 126, 123–130 (2015).
39. Cho, S.-Y., Chow, T.W.S., Leung, C.-T.: A neural-based crowd estimation by hybrid global learning algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 29, 535–541 (1999).
40. Chow, T.W.S., Cho, S.-Y.: Industrial neural vision system for underground railway station platform surveillance. Advanced Engineering Informatics. 16, 73–83 (2002).
41. Han, K., Wan, W., Yao, H., Hou, L.: Image Crowd Counting Using Convolutional Neural Network and Markov Random Field. arXiv:1706.03686 [cs]. (2017).
42. Zeng, L., Xu, X., Cai, B., Qiu, S., Zhang, T.: Multi-scale Convolutional Neural Networks for Crowd Counting. arXiv:1702.02359 [cs]. (2017).
43. Bertucco, L., Nunnari, G., Randieri, C., Rizza, V., Sacco, A.: A cellular neural network based system for cell counting in culture of biological cells. In: Proceedings of the 1998 IEEE International Conference on Control Applications (Cat. No.98CH36104). pp. 341–345 vol.1 (1998).
44. Biswas, D., Hongbo Su, Wang, C., Blankenship, J., Stevanovic, A.: An Automatic Car Counting System Using OverFeat Framework. Sensors (14248220). 17, 1–13 (2017).
45. Schofield, A.J., Mehta, P.A., Stonham, T.J.: A system for counting people in video images using neural networks to identify the background scene. Pattern Recognition. 29, 1421–1428 (1996).
46. Embleton, K.V., Gibson, C.E., Heaney, S.I.: Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. J Plankton Res. 25, 669–681 (2003).
47. Lal, S., Behera, S.K., Sethy, P.K., Rath, A.K.: Identification and counting of mature apple fruit based on BP feed forward neural network. In: 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS). pp. 361–368 (2017).
48. Nazeer, M., Wong, M.S., Nichol, J.E.: A new approach for the estimation of phytoplankton cell counts associated with algal blooms. Science of The Total Environment. 590–591, 125–138 (2017).
49. Taylor, J., Tisseyre, B., Bramley, R., Reid, A., Stafford, J.: A comparison of the spatial variability of vineyard yield in European and Australian production systems. Precision agriculture. 5, 907–914 (2005).
50. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision in C++ with the OpenCV Library. O'Reilly Media, Inc. (2013).

51. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M.: Tensorflow: a system for large-scale machine learning. In: OSDI. pp. 265–283 (2016).
52. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision. 88, 303–338 (2010).
53. Nam, H., Han, B.: Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4293–4302 (2016).
54. Chan, A.B., Liang, Z.-S.J., Vasconcelos, N.: Privacy preserving crowd monitoring: Counting people without people models or tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7 (2008).

Reading Between the Lines of Qualitative Data – How to Detect Hidden Structure Based on Codes

Alexander Keller[1], Hans Achatz[2]

[1] University of Passau, Chair of Information Systems (Information and IT-Service-Management), Passau, Germany
alexander.keller@uni-passau.de
[2] University of Passau, Teaching Unit Information Systems, Passau, Germany
achatz@uni-passau.de

Abstract. While qualitative research is experiencing broad acceptance in the information systems discipline, growing volumes of heterogeneous data pose challenges to manual qualitative analysis. We introduce an unsupervised machine learning approach based on graph partitioning to detect hidden information and structure in qualitative data samples. With the clustering technique, we map coded data to a graph and formulate a partitioning problem which is solved by integer linear programming. As a result, clusters of information sources are identified based on similarities given in the coded data. We demonstrate the approaches' ability to detect hidden information in coded qualitative data by application on coded interview transcripts. With the approach, we draw on a technique from the operations research discipline and expand the repertoire of approaches being used to analyze qualitative data in the context of information systems.

Keywords: qualitative data mining, clustering, graph partitioning, unsupervised learning, integer linear programming.

1 Introduction

Since 2005 one can see a significant growth in qualitative research publications in information system (IS) journals which shows the growing relevance of qualitative research in the IS discipline [1]. Besides studies that are applying qualitative methods to investigate research topics (e.g. [2–4]), a lot of work is focusing on guidelines regarding how qualitative research should be conducted within the IS discipline (e.g. [5–7]). Additionally, qualitative research is often seen as a minor discipline because of findings which are said to be accompanied with biases (e.g. subjectivity) regarding the common quality standards [8].

Qualitative research in IS aims to gather a deep understanding of behavioral and technical issues regarding the role of information technology (IT) and often supplements quantitative studies in a mixed method approach [9]. Referring to Romano Jr. et al. (2003) with an increasing amount of available qualitative data it becomes

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

necessary to develop approaches to analyze growing volumes [10]. In the context of qualitative data, especially clustering mechanisms have been discussed in literature (e.g. [11]) but still remain underused [12]. In the case of clustering being used, researchers typically perform manual techniques to detect groups based on codes (i.e. marker for a relevant information given in qualitative data) in qualitative data which requires extensive resource commitments [13]. Due to this there is a growing field of scholars who are applying machine-learning techniques on qualitative data. Within this field, automatic coding techniques based on natural language processing and graph theory (e.g. [14, 15]) as well as clustering approaches (e.g. [16]) were applied on qualitative data. However, the clustering techniques are mostly used to cluster participants with similar profiles of codes instead of clustering codes based on their relationship to each other. Following this research stream, the paper investigates the research question how hidden structure in qualitative data sets can be detected automatically based on code similarities? Answering this question, we develop an unsupervised learning technique called *CodeClust* based on the concept of graph partitioning. We draw on a technique which we developed to solve clustering problems in the operations research (OR) domain and adopt it to mine and analyze coded qualitative data. With performing the approach, results are generated that grant additional and hidden insights regarding the structure and affiliation of qualitative data constructs. With this, we intend to expand the repertoire of approaches that are used in qualitative IS research and contribute to future studies by providing a new approach to analyze qualitative data samples.

The remainder of paper is structured as follows: First, we give a brief overview of data collection, coding techniques and existing quality criteria in qualitative research. This leads to the third chapter in which we outline fundamentals and related work in the domain and present *CodeClust*. Besides describing data preparation and essential features of the approach, we illustrate how to use the technique on coded data from expert interviews. The paper concludes in a discussion of the approach and gives an outlook towards future research directions.

2 Data Collection, Coding Techniques and Essential Quality Criteria in Qualitative Research

There exist different techniques of data collection to gather empirical material for the purpose of data analysis. Non-numeric information is usually collected directly in form of interviews or indirectly from secondary sources like text documents [17]. This results in textual information sources like interview transcripts and case descriptions which serve as a foundation for further analysis. For a detailed view of different methods, data collection techniques, modes of analysis and quality criteria used in qualitative IS research, see Keller (2017) [18].

After the data has been collected, the information sources have to be structured for further analysis in a next step. Grounded theory is a widely used approach and serves as a foundation to structure textual information based on coding techniques [19]. The aim of the coding is to highlight segments of textual data that contain relevant information in regard to the underlying research question [20]. Therefore, codes represent information given in the data and serve as a generalization of information. In this field, Gläser and Laudel (2013) provide a detailed view on coding techniques and variations in objectives related to coding [19].

In most cases, the coding process itself is done manually. To ensure the quality of the results, the data sources are coded by different coders and their outputs are merged and evaluated for similarities. Within this coding process, software tools (e.g., *NVivo, Atlas.Ti, HyperResearch, MaxQDA*) are used by the coders and support in attaching codes to relevant text segments. Additionally, the tools provide functionalities in form of quantifications and visualizations to process the results.

In terms of quality criteria in qualitative research, validity and reliability are differently characterized than in the quantitative domain [21].
Although validity aims to ensure the quality and information value, Flick (2014) argues that flexibility is one major strength of qualitative approaches [22]. Therefore, communicative validation is a common approach to ensure credibility and accuracy in qualitative research [21, 23].
Reliability stands for the robustness of findings and the consistency of an approach. However, in qualitative research identical findings do not always represent reliable results (e.g. identical responses in interviews may point to prepared answers) [24]. In order to draw consistent conclusions in qualitative studies, the specific context and the data collection process itself must be described in very detail. With this, one can ensure traceability from an intersubjective point of view [25, 26].

In addition, credibility describes the internal validity of qualitative research which can be ensured by triangulation and negative case analysis [21]. As the researcher functions as a central part in the research process (i.e. data collection, coding, analysis, interpretation) some degree of bias is induced because of his personal perception which may lead to the problem of subjectivity [27, 28]. Therefore, qualitative research processes should be supplemented with standardized techniques to ensure a non-subjective evaluation and interpretation of qualitative data. Sarker et al. (2013) support this point of view by mentioning the strong "[…] need for clarity in the logic underlying data analysis […]" [29] in qualitative IS research.

3 Clustering Coded Qualitative Data

In qualitative research, scholars are interested in identifying behavioral or structural pattern to understand the studied phenomenon in the research context. For this purpose, the available heterogeneous qualitative data must be gathered and processed to structure the observed information. As mentioned in the last section, one central aspect of this systematic structuring process is the assignment of codes to specific information. While most of this processing is done manually in qualitative research, the machine learning (also referred as statistical learning) domain provides some techniques to predict and structure categorical information in an automated manner on the one side and offers approaches to scale up the coding process on the other side [30].

3.1 Foundations and Related Work

When predicting categories (i.e. discrete variables with nominal or ordinal scale) two different approaches can be distinguished due to the underlying structure of available data: (i) supervised learning describes techniques for classification of information based on labeled data. (ii) unsupervised learning techniques process unlabeled data to detect structure within the given information. Labeled data refers to a sample that has been tagged with labels and hence includes information about categories. Therefore, in supervised learning labeled training samples are used to train different classifiers (e.g. naïve bayes, support vector machines or decision trees) aiming to categorize information according to the training data. In contrast, unsupervised learning aims to detect information in form of clusters (e.g. k-means) or topics (e.g. latent dirichlet allocation) in unlabeled samples without any prior knowledge about categories. Figure 1 shows the difference between the two approaches of machine learning and gives some examples.

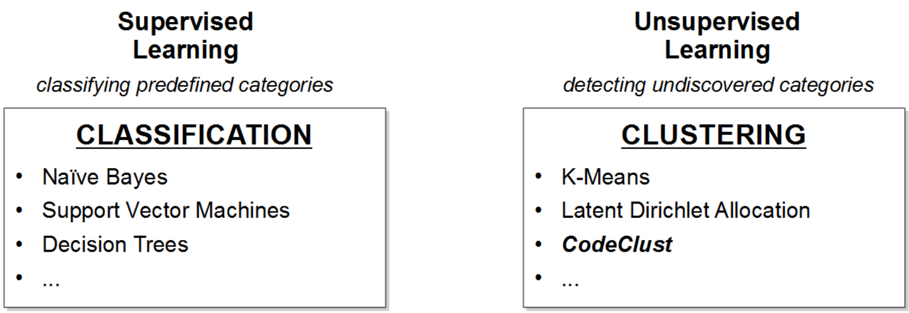


Figure 1. Unsupervised vs. Supervised Prediction of Categories.

As we aim to detect hidden information in qualitative data samples, this paper focuses on unsupervised learning in form of clustering. Clusters are defined as sets of objects (e.g. text documents) which are grouped based on similarities. Most prior work in this field has its origin in text mining, where documents are clustered based on the similarity of words to identify certain semantic topics [31, 32]. Besides emphasizing content representation, clustering is also used to identify dominant information in given samples [33].

The foundation of previous research is in general based on partitional clustering methods like K-Means [34] or graph-partitioning approaches [35] where an object belongs in exactly one cluster and the number of clusters is given in advance. Besides that, hierarchical clustering techniques are used where a set of nested clusters is built by successive merging or splitting [36, 37]. In general, hierarchical clustering strategies fall into two types: (i) the agglomerative or bottom up approach where clusters are merged on every hierarchy level and (ii) the divisive or top down approach in which splits are performed on every hierarchy level. Another approach is called topic modeling which decomposes information sources into topics and links the information sources to the identified topics based on certain probabilities. Latent dirichlet allocation introduced by Blei et al. (2003) is a common technique for topic modeling and is applicable in most textual based clustering scenarios [39, 40].

Most of the previous work has in common that the occurrence, frequency and/or combination of words are taken into consideration when building clusters. However, in qualitative research additional information in form of codes is available which can be considered in the clustering to detect hidden information. With this paper, we propose a clustering technique that grasps the relation of codes to each other.

3.2 Concept

Meeting the needs for an automated clustering technique based on the coded qualitative data, we develop a graph-partitioning approach called *CodeClust.* The presented technique is built upon existing clustering approaches and embraces the numerical structure of coded data which results from the assignment of text segments to codes (e.g. coding process in grounded theory and content analysis). The intention of the approach lies in clustering information sources and codes. As Marton (2013) points out, this is necessary because "[…] collected data needs to be grouped […] in order to form relevant corpora for comparison" [41]. Such additional information about groups serves as a basis and guidance for detailed interpretation and supports the analysis of subgroups [16, 42].

In our scenario, we assume that these similarities are based on information sources (e.g. interview transcripts) expressing the same ideas and therefore standing in relation to each other. For example, groups of experts can be identified in interviews which are considered to be similar relating to their statements. As *CodeClust* is a partitional clustering each object is assigned to one cluster.

The provided graph-theoretic approach has originally been developed to simplify the complexity of staff scheduling problems in the OR discipline [43]. We adopt our method from this domain and adapt it to analyze qualitative data to cluster affiliated information sources in groups with regard to their coding. Before describing the approach table 1 defines the used mathematical notations.

Table 1. Mathematical Notations and Definition

Notation	Definition
n	Total number of information sources
m	Total number of codes
i	Index of information sources, $i \in \{1, \ldots, n\}$
V	Set of nodes within a graph, with $\|V\|$ being the total number of nodes
E	Set of edges between nodes of V within a graph
c	Index of codes, $c \in \{1, \ldots, m\}$
f_{ci}	Code-frequency of code c within information source i, $f_{ci} \in \mathbb{N}_0$
b_{ci}	Binary value for a code being coded above average within i

In a first step, the coded qualitative data must be represented as a graph to use the technique. Therefore, the codes c are represented as nodes V in a graph $G = (V, E)$. The total number of nodes $|V|$ is m, representing the total number of codes given in the data set. E represents a set of edges between nodes c and c' of V. An edge between two nodes exists, if both codes represented by the nodes are coded above average within at least one information source. This is modeled with the binary value b_{ci} that takes the following two states:

$$b_{ci} = \begin{cases} 1, if\ f_{ci} > \left[\frac{1}{|V|}\sum_{c'=1}^{|V|} f_{c'i}\right] \\ 0, otherwise \end{cases}$$

These binary values b_{ci} determine the graph, which serves as a basis for the partitioning technique. In addition, each edge of the graph contains a weight $g_{cc'}$ that stands for the number of information sources in which the two codes c and c' are coded above average:

$$g_{cc'} = \sum_{i=1}^{n} b_{ci}\, b_{c'i}$$

The graph partitioning problem is solved with integer linear programming (ILP) to identify two disjunctive sets of nodes $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. For the ILP two binary auxiliary variables are introduced:

$$z_c = \begin{cases} 1, if\ the\ first\ set\ of\ nodes\ contains\ c \\ 0, otherwise \end{cases}$$

$$y_{cc'} = \begin{cases} 1, if\ an\ edge\ between\ c\ and\ c'\ leads\ from\ one\ subgraph\ to\ the\ other \\ 0, otherwise \end{cases}$$

The partitioning aims to minimize the weights of edges which lead from subgraph G_1 to G_2. To solve this minimization problem, the objective function is formulated under four constraints:

$$\min \sum_{(c,c')\in E} g_{cc'}\, y_{cc'}$$

$$(1) \qquad z_c - z_{c'} - y_{cc'} \leq 0 \qquad \forall (c,c') \in E$$

$$(2) \qquad z_{c'} - z_c - y_{cc'} \leq 0 \qquad \forall (c,c') \in E$$

$$(3) \qquad \sum_{c\,\in\,V} z_c \geq \frac{|V|}{4}$$

$$(4) \qquad \sum_{c\,\in\,V} z_c \leq \frac{3|V|}{4}$$

Constraints 1 and 2 ensure that the weights of edges are considered in the objective function if the particular edge leads from one subgraph into the other. E.g. the first and second constraint forces $y_{cc'}$ to be 1 when c and c' are nodes of different subsets (e.g. $z_c = 1$; $z_{c'} = 0$). Hence the weight of the related edge between c and c' $g_{cc'}$ is considered in the objective function. In the approach, the constraints 3 and 4 are used to prevent the clusters from becoming very small or vice versa very large. With the actual setting, it is ensured that a cluster contains at least the fourth part of the total amount of nodes. These boundaries can be set according to the individual preferences on the minimum cluster size.

In figure 2, an example is given which shows a graph G holding six nodes with different weights on edges between the nodes. After solving the ILP for the given example two subgraphs $G_1$and G_2 could be identified where two edges with a weight of 1 are leading from subgraph G_1 to the subgraph G_2 and vice versa. In the example the minimal solution of the objective function results in $g_{2\,3} + g_{1\,5} = 2$.

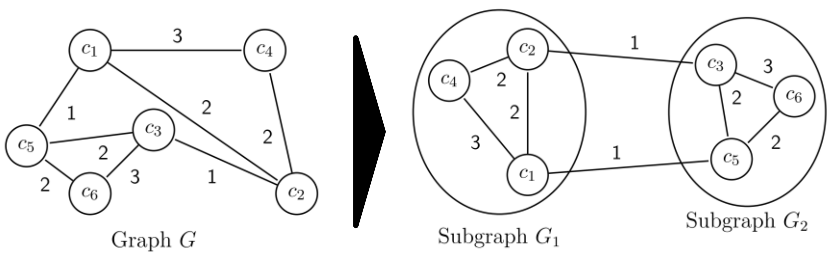


Figure 2. Example of Graph Partitioning

As shown in Achatz (1999) the technique can also be performed multiple times on large datasets to generate more than two subgraphs [43]. As multiple iterations can be performed by using the output from one iteration as input for the next one, large qualitative data samples can be analyzed with the approach in an automated manner. In settings where a code is not mentioned above average, one can either (i) exclude the code from the partitioning, (ii) assign code to smaller partition or (iii) assign code to larger partition. In scenarios where every node must be assigned to a cluster option (ii) or (iii) would be applied. However, in terms of clustering codes, we recommend using the first option. The specific code is excluded from the partitioning. This does not mean that the code should be excluded from further analysis. The information from the general coding can still be valuable for later interpretation.
As in each iteration the sample is only split into two clusters, the appropriate amount of clusters can be determined with existing measures like the elbow curve method [44] or a silhouette analysis [45]. In the first method the percentage of variance explained is used to find the appropriate number of clusters. The second method provides a graphical representation in form of silhouettes and measures how well an object fits to its cluster compared to other clusters. Identifying the correct number of clusters, however, is not a trivial task and there exists a multitude of cluster quality indicators (e.g. [46, 47]).

Regarding the behavior of the approach in higher orders of magnitude, there does not exist a high sensitivity to the number of observations. However, larger sample sizes will increase the runtime to solve the formulated partitioning problem with ILP.

3.3 Interpretation

The results from the clustering technique help to answer the question which set of information sources mention which specific codes. Hence, groups can be identified that include information sources based on the similarity of their coding. The clustering serves as an additional insight and makes it possible to identify structure in form of dependencies between information sources and codes. By just considering the codes, which are mentioned above average within the information source, the method uses the code-frequency for clustering. Therefore, it is assumed that information sources that mention the same codes above average are related to each other. This relation results from the fact that the information sources contain similar content regarding the research objective. E.g. if the information sources are interviews one can build groups of experts based on the similarity of their content-related statements. As a result, it can be identified which group of experts emphasizes which codes. Combining this with additional information about experts one can generate specific insights regarding the meaning and affiliation of identified codes.

4 Application to Field Data

The approach has been used on data from interviews with entrepreneurship experts to identify success indicators for IT-start-ups [48]. Besides entrepreneurs and investors, business angels were chosen as experts to generate a holistic view on the topic. In total eleven interview transcripts were coded on the basis of the methodology proposed from Steigleder (2008) [49]. The theory orientated content coding technique results in 22 codes. Each code represents a separate success indicator in the domain. The data set in table 2 represents the results of the coding process of textual data from the interviews.

Table 2. Data Set of Coded Qualitative Data from Interviews

			Interview i										
	c	Success Indicator	1	2	3	4	5	6	7	8	9	10	11
Code c	1	Sales competence and marketing power	2	3	5	2	0	2	1	2	3	0	1
	2	Perseverance	0	0	0	1	2	2	1	1	0	0	0
	3	Value orientated thinking	4	1	0	0	2	2	0	0	0	0	0
	4	Entrepreneurship and professional experience	2	1	1	2	2	2	2	2	0	0	2
	5	Industry specific competence	2	0	3	3	1	0	1	3	1	0	0
	6	ICT competence	1	2	0	2	1	1	1	4	2	0	0
	7	Conversion capability and speed	3	0	2	1	0	0	0	0	1	0	2
	8	Staff management skills	2	1	0	1	1	0	0	0	1	0	3
	9	Scalability and market ability	2	2	0	1	0	2	3	1	3	0	0
	10	Proof of feasibility and verification	0	0	1	0	1	4	0	1	0	2	0
	11	Business model flexibility and independence	2	3	0	0	2	0	0	2	1	1	2
	12	Team composition	1	1	1	2	2	0	0	2	2	3	0
	13	Industry specific financing	1	0	1	7	0	1	1	6	0	2	0
	14	Accelerator or incubator program	0	0	7	0	0	0	0	0	1	2	0
	15	R&D cooperations	0	0	0	3	1	4	2	2	0	0	1
	16	Seed-customer and technology partner	1	5	0	0	5	1	1	2	1	0	1
	17	Handling of market conditions	0	7	0	1	4	1	4	5	1	0	0
	18	Political and regulatory business environment	0	1	0	3	5	0	0	2	3	0	1
	19	Industry specific norms and requirements	0	4	0	3	3	0	1	2	3	0	0
	20	Customer orientated problem solving	0	1	3	1	3	3	0	2	2	2	1
	21	Feedback driven product development	0	2	1	0	3	0	0	5	3	1	1
	22	Prototype orientated product development	1	0	0	2	7	0	1	1	1	2	1

The values given in the matrix stand for the code-frequency of a particular code in an information source (f_{ci}). E.g. code *c=1* is coded two times in interview *i=1* which results in f_{11} equals 2. The variable b_{ci} is a binary value which equals 1 if a code *c* is coded above average in an information source *i*. If this is not the case, b_{ci} is set to 0.

Before performing the approach, the graph must be generated based on the values of b_{ci}. An edge between two nodes (i.e. codes) *c* and *c'* exists, if for at least one information source i the multiplication of b_{ic} and $b_{ic'}$ equals 1. The number of information sources i for which this is true represents the weight $g_{cc'}$ of a particular edge between the nodes *c* and *c'*. As represented in figure 3 on the next page the three codes (*c*=10; *c*=15; *c*=20) are mentioned above average in the sixth interview (i=6) which results in $b_{10\,6} = b_{15\,6} = b_{20\,6} = 1$. Besides that, information sources 6 and 10 mention the codes 10 and 20 above average. Therefore, the weight of $g_{20\,10}$ equals 2.

In figure 3 the input and the output of the approach is visualized. The matrix on the left side shows an unsorted data set of binary values b_{ci} based on the qualitative data given in table 2. This represents the graph which serves as an input for the presented approach. The matrix on the right side in figure 3 represents the data set sorted into two groups based on the graph partitioning. The first group contains the sources 1, 3, 6 and 10 as well as the codes 20, 12, 10, 14 and 3. The second group is determined by the rest of sources and codes, while code 2 is an artefact that cannot be assigned to any group because it is not mentioned above average in any information source. One can see that no source of the second group mentions a code of the first group above average, which means that each code of the first group is uniquely assigned to it.

	Interview i										
c	1	2	3	4	5	6	7	8	9	10	11
1	0	**1**	**1**	0	0	0	0	0	**1**	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	**1**	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	**1**	0	0	0	**1**
5	0	0	**1**	**1**	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	**1**	0	0	0
7	**1**	0	0	0	0	0	0	0	0	0	**1**
8	0	0	0	0	0	0	0	0	0	0	**1**
9	0	0	0	0	0	0	**1**	0	**1**	0	0
10	0	0	0	0	0	**1**	0	0	0	**1**	0
11	0	**1**	0	0	0	0	0	0	0	0	**1**
12	0	0	0	0	0	0	0	0	0	**1**	0
13	0	0	0	**1**	0	0	0	**1**	0	**1**	0
14	0	0	**1**	0	0	0	0	0	0	**1**	0
15	0	0	0	**1**	0	**1**	**1**	0	0	0	0
16	0	**1**	0	0	**1**	0	0	0	0	0	0
17	0	**1**	0	0	**1**	0	**1**	**1**	0	0	0
18	0	0	0	**1**	**1**	0	0	0	**1**	0	0
19	0	**1**	0	**1**	0	0	0	0	**1**	0	0
20	0	0	**1**	0	0	**1**	0	0	0	**1**	0
21	0	0	0	0	0	0	0	**1**	**1**	0	0
22	0	0	0	0	**1**	0	0	0	0	**1**	0

	Interview i										
c	1	3	6	10	2	4	5	7	8	9	11
20	0	1	1	1	0	0	0	0	0	0	0
12	0	0	0	1	0	0	0	0	0	0	0
10	0	0	1	1	0	0	0	0	0	0	0
14	0	1	0	1	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0
1	0	**1**	0	0	**1**	0	0	0	0	**1**	0
16	0	0	0	0	**1**	0	**1**	0	0	0	0
5	0	**1**	0	0	0	**1**	0	0	0	0	0
21	0	0	0	0	0	0	0	0	**1**	**1**	0
22	0	0	0	**1**	0	0	**1**	0	0	0	0
9	0	0	0	0	0	0	0	**1**	0	**1**	0
18	0	0	0	0	0	**1**	**1**	0	0	**1**	0
17	0	0	0	0	**1**	0	**1**	**1**	**1**	0	0
8	0	0	0	0	0	0	0	0	0	0	**1**
13	0	0	0	**1**	0	**1**	0	0	**1**	0	0
4	0	0	0	0	0	0	0	**1**	0	0	**1**
6	0	0	0	0	0	0	0	0	**1**	0	0
7	**1**	0	0	0	0	0	0	0	0	0	**1**
19	0	0	0	0	**1**	**1**	0	0	0	**1**	0
11	0	0	0	0	**1**	0	0	0	0	0	**1**
15	0	0	**1**	0	0	**1**	0	**1**	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0

Group 1 Group 2

Figure 3. Results Generated by the Clustering Technique

Using the approach, we identified that in the first group experts are clustered which have a strong relation to phases at the beginning of the start-up life cycle. In this group is one business angel supporting start-ups in the seed and start-up phase and the rest of experts are entrepreneurs standing at the beginning of their ventures. This observation is accompanied with the fact that the codes of the first group represent success indicators which are important when founding and developing the firm. E.g. *team composition* (c=12) and *accelerator or incubator program* (c=14) are aspects entrepreneurs must deal with at the very start of a venture.

In contrast, experts in the second cluster handle start-up firms in the growth and expansion phase (e.g. investors, advanced founders). Looking at the codes it becomes evident that the success indicators *handling of market conditions* (c=17) and *political and regulatory business environment* (c=18) are related to advanced phases in the start-up lifecycle as well. Additionally, no expert of group two is mentioning success indicators of group one above average. In contrast, some experts of group one do mention codes of group two. This shows that no expert in group two (i.e. expert related to advanced start-up life cycle phases) emphasizes the success indicators of early start-up phases because they seem to be less relevant. Experts in group one (i.e. experts related to early start-up life cycle phases) consider the future by focusing not only on success indicators for the beginning of a venture but also on success indicators of advanced phases of the start-up lifecycle.

While choosing start-up stakeholders as experts we did not focus on their affiliation regarding different start-up lifecycle phases. With performing the clustering, we are able to detect hidden structure in form of subgroups which generates valuable findings concerning the affiliation of success indicators to lifecycle phases.

5 Limitations and Future Research

As shown in the example the approach provides valuable results and assists in the detection of hidden structure in qualitative data samples. However, the technique is subject to some limitations.

First, it is depending on the underlying data. This means, that the results generated can only be as good as the quality of the gathered qualitative data. Therefore, the clustering should only be performed on top of a robust data collection and coding process.

Second, although it is transparent how clusters are generated, the findings might still be biased by a subjectivity in data pre-processing steps and by a wrong interpretation of clustering results. Therefore, it is important to mention that the clustering results offer guidance for interpretation and subgroup analysis but should not be applied as strict decision guidelines.

Third, like other techniques analyzing qualitative data, the approach provides an indication for the affiliation of codes with regard to the underlying research subject. In contrast to "[…] statistical methods which require representative data, cluster analysis does not find generalizable characteristics" [12]. Hence, it is suitable for qualitative research which aims to understand complex phenomena regarding a specific topic of interest.

Fourth, although the used code-frequency is a suitable measure to create the graph of codes it can make sense to include other aspects like the context a code is related to or actual speech such as laughter. Hence, the frequency measure could be enhanced or replaced with other subject related aspects.

Fifth, like in other clustering techniques, it has to be considered that there may be settings where the clustering does not result in interpretable results. This is indicated by a high value for the objective function related to the sum of all weights of edges. In this case, there exist outliers which indicate an unstable clustering. Therefore, we recommend combining multiple manual and automated clustering techniques to confirm accuracy of generated cluster solutions.

Regarding the limitations, future research should investigate how the approach can be combined with already existing qualitative data analysis techniques. Especially the possibilities for complementation should be considered to validate findings on the one side and supplement them on the other. In addition, the behavior of the presented approach could be studied in more detail. E.g. the clustering technique can be tested with a variation of the underlying data. With this the behavior of the partitioning

procedure can be evaluated with focus on the robustness of the clustering. Furthermore, we intentionally applied a small data example to introduce the approach in a simple and understanding manner. Although the clustering delivers stable results for larger data samples in its original discipline [43], further research should investigate the behavior with different dimensions of volume. A benchmark of the approach against other techniques can be based on agreement measures like the Cohen's kappa [50, 51]. In a next step, the technique could also be compared to traditional connectivity-based clustering techniques like hierarchical clustering. As distance metric the proposed $g_{cc'}$ will be considered to identify similarity of codes.
In addition, the approach is not only suitable in qualitative research scenarios but should also be considered in the domain of business analytics. Although in these scenarios the data modeling might be different, the problem formulation remains the same.

6 Conclusion

We introduced a machine learning approach to detect hidden information and structure in qualitative data samples. Therefore, we used a clustering technique based on a graph theory to group information sources based on the similarity of codes. With the approach, we map the coded data to a graph and formulate a graph partitioning problem which is solved with ILP. As a result, the technique separates the graph into different clusters based on the similarities given in the coded data. The technique is designed to be used on textual qualitative data which results from any kind of qualitative coding process. Hence, the approach can be performed independently of the underlying methodology and does not replace any existing procedures but complements them.

Until now interpretation of qualitative data mostly relies on the system of codes, their assigned textual passages and different frequency measures. Most clustering in this context is performed manually or semi-automated which is contradictory to the essential quality criteria of intersubjective traceability [22, 29]. Hence, with the presented automated approach we present a new data modeling and clustering technique which adds to the existing repertoire of qualitative data analysis methods in IS. Although the data pre-processing (i.e. data collection, transcription and coding process) might still be biased by subjectivity, the approach could increase the reliability of qualitative research approaches regarding interpretation. However, the clustering should guide and support the researcher in context informed interpretation and code relationship analysis but should not be used in form of decision rules to split qualitative data sets.
Referring to Sarker et al. (2012) about 60% of qualitative studies in the IS discipline use coding procedures to analyze empirical data [1]. Because of this, our approach could address and complement many existing and future qualitative studies in the IS domain in terms of applying a mixed methods approach to answer research questions. Especially when dealing with large volumes of empirical data, the potential of the clustering could be exploited.

References

1. Sarker, S., Xiao, X., Tanya, B.: Towards an Anatomy of "Successful" Qualitiative Research Manuscripts in IS: A Critical Review and Some Recommendations. In: Thirty Third International Conference on Information Systems (ICIS 2012). pp. 1–21 (2012).
2. Chua, C.E.H., Yeow, A.Y.K.: Artifacts, Actors, and Interactions in the Cross-project Coordination Practices of Open-source Communities. J. Assoc. Inf. Syst. 11, 838–867 (2010).
3. Davidson, E.J., Chismar, W.G.: The Interaction of Institutionally Triggered and Technology - Triggered Social Structure Change: An Investigation of Computerized Physician Order Entry. MIS Q. 31, 739–758 (2007).
4. Vogelsang, K., Steinhüser, M., Hoppe, U.: A Qualitative Approach to Examine Technology Acceptance. In: International Conference on Information Systems (ICIS 2013). pp. 234–245 (2013).
5. Urquhart, C., Lehmann, H., Myers, M.D.: Putting the "Theory" Back into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. Inf. Syst. J. 20, 357–381 (2010).
6. Klein, H.K., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Q. 23, 67–93 (1999).
7. Schultze, U.: A Confessional Account of an Ethnography about Knowledge Work. MIS Q. 24, 3–41 (2000).
8. Conboy, K., Fitzgerald, G., Mathiassen, L.: Qualitative Methods Research in Information Systems: Motivations, Themes, and Contributions. Eur. J. Inf. Syst. 21, 113–118 (2012).
9. Venkatesh, V., Brown, S.A., Bala, H.: Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. Manag. Inf. Syst. Q. 37, 21–54 (2013).
10. Romano Jr., N.C., Donovan, C., Chen, H., Nunamaker Jr., J.F.: A Methodology for Analyzing Web-Based Qualitative Data. J. Manag. Inf. Syst. 19, 213–246 (2003).
11. Guest, G., McLellan, E.: Distinguishing the Trees from the Forest: Applying Cluster Analysis to Thematic Qualitative Data. Field methods. 15, 186–201 (2003).
12. Macia, L.: Using Clustering as a Tool: Mixed Methods in Qualitative Data Analysis. Qual. Rep. 20, 1083–1094 (2015).
13. Brickey, J., Walczak, S., Burgess, T.: A Comparative Analysis of Persona Clustering Methods A Comparative Analysis of Persona Clustering Methods. (2010).
14. Crowston, K., Liu, X., Allen, E.E.: Machine learning and rule-based automated coding of qualitative data. In: Proceedings of the ASIST Annual Meeting. pp. 1–2 (2010).
15. Tierney, P.: A Qualitative Analysis Framework Using Natural Language Processing and Graph Theory. Int. Rev. Res. Open Distrib. Learn. 13, 173 (2012).
16. Henry, D., Dymnicki, A.B., Mohatt, N., Allen, J., Kelly, J.G.: Clustering Methods with Qualitative Data: a Mixed-Methods Approach for Prevention Research with Small Samples. Prev. Sci. 1007–1016 (2015).
17. Myers, M.D.: Qualitative Research in Information Systems. MIS Q. 21, 241–242 (1997).
18. Keller, A.: How to Gauge the Relevance of Codes in Qualitative Data Analysis ? – A Technique Based on Information Retrieval. In: 13th International Conference on Wirtschaftsinformatik. pp. 1096–1110 (2017).
19. Gläser, J., Laudel, G.: Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations. Forum Qual. Sozialforsch. 14, (2013).

20. Kelle, U.: Theory Building in Qualitative Research and Computer Programs for the Management of Textual Data. Sociol. Res. 2, (1997).
21. Lincoln, Y.Y., Guba, E.G.: Naturalistic Inquiry. Sage, Beverly Hills (1985).
22. Flick, U.: Gütekriterien qualitativer Sozialforschung. In: Baur, N. and Blasius, J. (eds.) Handbuch Methoden der empirischen Sozialforschung. pp. 411–423. Springer, Wiesbaden (2014).
23. Bryman, A.: Research Methods and Organization Studies. Unwin Hyman, London (1989).
24. Flick, U.: An Introduction to Qualitative Research. Sage, Los Angeles (2014).
25. Kruse, J.: Qualitative Interviewforschung: Ein integrativer Ansatz. Beltz Juventa, Weinheim (2014).
26. Steinke, I.: Kriterien qualitativer Forschung. Beltz Juventa, Weinheim (1999).
27. Zahedi, F.M., Van Pelt, W.: Web Documents' Cultural and Femininity Masculinity. J. Manag. Inf. Syst. 23, 87–128 (2006).
28. Wrona, T., Gunnesch, M.: The One Who Sees More is More Right: How Theory Enhances the 'Repertoire to Interpret' in Qualitative Case Study Research. J. Bus. Econ. 1–27 (2015).
29. Sarker, S., Xiao, X., Beaulieu, T.: Guest Editorial: Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles. MIS Q. 37, iii–xviii (2013).
30. Chen, N., Kocielnik, R., Drouhard, M., Peña-Araya, V.: Challenges of Applying Machine Learning to Qualitative Coding. In: ACM SIGCHI Workshop on Human-Centered Machine Learning (2016).
31. Pons-Porrata, A., Berlanga-Llavori, R., Ruiz-Shulcloper, J.: Topic Discovery Based on Text Mining Techniques. Inf. Process. Manag. 43, 752–768 (2007).
32. Gupta, V., Lehal, G.S.: A Survey of Text Summarization Extractive Techniques. J. Emerg. Technol. Web Intell. 2, 258–268 (2010).
33. Guo, X., Wei, Q., Chen, G., Zhang, J., Qiao, D.: Extracting Representative Information on Intra-Organizational Blogging Platforms. MIS Q. 41, 1105–1127 (2017).
34. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. ACM Comput. Surv. 31, 264–323 (1999).
35. Aggarwal, C.C., Wang, H.: A Survey of Clustering Algorithms for Graph Data. In: Managing and Mining Graph Data. Advances in Database Systems. pp. 275–301. Springer, Boston (2010).
36. Berkhin, P.: A Survey of Clustering Data Mining Techniques. Group. Multidimens. Data. 25–71 (2006).
37. Lee, C.-J., Hsu, C.-C., Chen, D.-R.: A Hierarchical Document Clustering Approach with Frequent Itemsets. Int. J. Eng. Technol. 9, 174–178 (2017).
38. Blei, D.M., Edu, B.B., Ng, A.Y., Edu, A.S., Jordan, M.I., Edu, J.B.: Latent Dirichlet Allocation. J. Mach. Learn. Res. 3, 993–1022 (2003).
39. Gefen, D., Endicott, J.E., Fresneda, J.E., Miller, J., Larsen, K.R.: A Guide to Text Analysis with Latent Semantic Analysis in R with Annotated Code: Studying Online Reviews and the Stack Exchange Community. Commun. Assoc. Inf. Syst. 41, 450–496 (2017).
40. Debortoli, S., Müller, O., Junglas, I., vom Brocke, J.: Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. Commun. Assoc. Inf. Syst. 39, 110–135 (2016).
41. Marton, A.: Purposive Selection and the Quality of Qualitative IS Research. In: Thirty Fourth International Conference on Information Systems (ICIS 2013) (2013).
42. Renaud, A., Walsh, I., Kalika, M.: Is SAM Still Alive? A Bibliometric and Interpretive Mapping of the Strategic Alignment Research Field. J. Strateg. Inf. Syst. 25, 75–103 (2016).

43. Achatz, H.: Partitionierung von Werkern aufgrund von Qualifikationsprofilen. In: Kossbiel, H. (ed.) Modellgestützte Personalentscheidungen. pp. 143–158. Rainer Hampp, München (1999).
44. Thorndike, R.L.: Who Belongs in the Family? Psychometrika. 18, 267–276 (1953).
45. Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. J. Comput. Appl. Math. 20, 53–65 (1987).
46. Ackerman, M., Ben-David, S.: Measures of Clustering Quality: A Working Set of Axioms for Clustering. Adv. Neural Inf. Process. Syst. 121–128 (2009).
47. Nguyen, Q.H., Rayward, Smith, V.J.: Internal Quality Measures for Clustering in Metric Spaces. Int. J. Bus. Intell. Data Min. 3, 4–29 (2008).
48. Keller, A.: Zum Erfolg von IKT-Start-ups in der deutschen Elektrizit tswirtschaft - Eine explorative Studie auf Basis von Experteninterviews. Dr. Kova , Hamburg (2016).
49. Steigleder, S.: Die strukturierende qualitative Inhaltsanalyse im Praxistest: eine konstruktiv kritische Studie zur Auswertungsmethodik von Philipp Mayring. Tectum-Verlag, Marburg (2008).
50. Reilly, C., Wang, C., Rutherford, M.: A Rapid Method for the Comparison of Cluster Analyses. Stat. Sin. 15, 19–33 (2005).
51. Fraley, C., Raftery, A.E.: How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. Comput. J. 41, 578–588 (1998).

Online Auctions with Dual-Threshold Algorithms: An Experimental Study and Practical Evaluation

Philipp Knöpfle[1], Johannes Knörr[1], Sören Merting[2]

[1] Technical University of Munich, Department of Informatics, Munich, Germany
{philipp.knoepfle,johannes.knoerr}@tum.de
[2] Technical University of Munich, Department of Informatics, Munich, Germany
{soeren.merting}@in.tum.de

Abstract. Online auctions are a viable alternative to conventional posted price mechanisms. Agrawal, Wang, and Ye [1] have proposed two primal-dual algorithms for revenue-maximizing multi-item allocation tasks. Although promising in terms of theoretical properties and competitive ratios, there is a lack of evidence regarding the real-world practicability of these mechanisms, for instance referring to online auction-based tickets sales. In this paper, we conduct an experimental study on both the *One-Time Learning Algorithm* (OLA) and the *Dynamic Learning Algorithm* (DLA) based on synthetic data, revealing the remarkable aptitude of the latter for non-trivial online auctions. Being robust to most input variations, the inherent dynamic update of dual thresholds achieves a superior balance with respect to the trade-off between objective function values and runtimes. We address critical sensitivities quantitatively and draft several small extensions by incorporating input distribution knowledge.

Keywords: Online Auctions, Online Ticket Sales, Experimental Study

1 Introduction

Online auctions and auction-type mechanisms become increasingly popular for revenue-maximizing allocations of scarce resources. While display ad auctions and sponsored search are always based on bidders revealing their willingness-to-pay prior to the actual allocation, many B2C business models still rely on conventional posted prices. For instance, ticket sales for cultural events or sports competitions are often conducted with a fixed-price policy, occasionally replaced by concepts of revenue management or dynamic pricing. In this case, however, sellers have little knowledge on the willingness-to-pay or consumer surplus of the bidders, potentially preventing higher revenues. In contrast, auction-type allocations in an online fashion can enable a better absorption of buying power and hence a more efficient allocation. Recently, in the aviation or hotel industry it can be observed that companies experiment with their established pricing models by gradually substituting conventional posted price

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

mechanisms with online open auctions. Such auctions might also be of particular interest for ticket sales in order to reduce black market activity. Aside from the online arrival of bidders, however, ticket allocation problems often exhibit additional complexity as a result of excess demand, heterogeneous willingness-to-pay and short processing times, calling for performant and fast decision-making algorithms.

We examine a notable contribution by [1] in the context of online resource allocation. Based on a primal-dual approach, the authors state theoretical properties and postulate a broad applicability of their two algorithms vis-a-vis auction-type allocations. As with many primal-dual frameworks, however, little is known about the implementation, real-world feasibility or practical challenges of these mechanisms.

This paper seeks to examine this gap between theory and practicability by assessing the empirical applicability of both the *One-Time Learning Algorithm* (OLA), which calculates a single set of dual threshold prices, and the *Dynamic Learning Algorithm* (DLA), which continuously updates dual prices at geometric time intervals. Optimizing the revenue against a stationary bidding process, we perform numerical experiments on both algorithms based on synthetic data and compare the runtimes and objective values to three benchmarks. Furthermore, we examine reasonable parameter combinations and investigate the algorithms' sensitivity with respect to their input parameters. We find a strong trade-off between outcome quality and runtime, whereas the DLA produces superior outputs while maintaining moderate runtimes for almost all cases. Our experiments provide evidence that the DLA is especially suited for complex allocation tasks with resource scarcity, heterogeneous bidders and limited ex-ante knowledge. Moreover, addressing the most critical sensitivity, we suggest several extensions utilizing priorly known input distribution information and thus improving the practical applicability for online auctions.

The remainder of this paper is structured as follows. Section 2 begins by reviewing recent contributions in the online resource allocation literature. We also explain the algorithms of [1] in more detail. We present our implementation and experimental design in section 3.1. Section 3.2 provides the experimental observations and states key results. We discuss our findings with respect to extensions and put emphasis on enhancing the practicability. Section 4 concludes and suggests further research.

2 Review of Literature and Research

Optimizing the allocation of given resources with respect to revenue or social welfare is a core element of a variety of scientific contributions. A popular approach is to actively control the availability of resources, often referred to as revenue management [16]. Closely related is the concept of dynamic pricing, in particular dynamically adjusting posted prices for the purpose of revenue maximization. Comprehensive literature overviews for value-based pricing techniques are, for instance, provided by [12] or [16]. Although these concepts ultimately aim at exploiting differences in customer willingness-to-pay, one key drawback of such techniques is the failure to explicitly record willingness-to-pay and hereby enhance the producer surplus.

Another domain of resource allocation is auction theory. While other objectives such as efficiency or states of equilibria can certainly exist, revenue maximizations might also be pursued through auctions. Since bidders reveal their true willingness-to-pay in incentive-compatible mechanisms, sellers might be able to absorb a larger proportion of the consumer surplus, thus increasing revenue. Traditional auction theory usually refers to a single point in time with comprehensive knowledge of demand and supply. However, an increasing number of digital markets are dynamic, meaning that bidders reveal their willingness-to-pay sequentially, requiring an immediate allocation decision by the auctioneer. The growing field of online auctions focuses – amongst others – on the algorithmic design for this class of problems. Examples include, for instance, [2], [14], and [17]. [13] specifically address the case of online ticket sales and design a fully stochastic and dynamic algorithm in order to compute an optimal online auction mechanism. [9] demonstrate that posted price mechanisms can indeed match competitive ratios of combinatorial auction principles, provided that stochastic information of the bidders' valuations is available.

Among a variety of algorithmic frameworks, one popular approach for online allocations is the utilization of dual threshold prices. Reasons for this include, for instance, a wide field of application, simple interpretability as well as the ease of implementation. In essence, this concept makes use of shadow prices of an ex-ante linear program as weights for subsequent decision-making. An incoming request can thus only be accepted if the weighted resource consumption is exceeded by the communicated willingness-to-pay. For instance, [8] investigate the AdWords problem within a similar framework as [1]. Under a random permutation assumption with known number of bidders and a specific right-hand side condition, they retrieve a $(1-\varepsilon)$-competitive algorithm. In a similar manner, but assuming an i.i.d. input with unknown and changing distributions, [7] develops dual-based resource allocation algorithms for stochastic AdWords problems. In a series of interrelated contributions, [3-5] use primal-dual approaches to match and prove several competitive ratios for various kinds of online problems, for example ad-auctions. Other examples using dual approaches include [10-11] for display ads or [6] in a Bayesian auction setting.

However, some contributions, for example those relating to the AdWords problem, do not permit multidimensional demand vectors, in particular the possibility to request several resources simultaneously, and are thus not directly applicable to ticket sales or similar B2C businesses. In contrast, the algorithms proposed by [1] explicitly refer to classical multi-item revenue maximization problems. The authors also demonstrate a superiority with respect to the theoretical competitive ratios compared to related online auction frameworks.

In order to allocate resources in an online auction fashion [1] use online linear programming, where the constraint matrix and corresponding objective function coefficients are revealed column by column over time. The linear program is calculated based on the input received so far and without any information about future requests. However, all subsequent decisions are based on this solution. More precisely, consider the following linear program:

$$\begin{aligned} &\text{maximize } \textstyle\sum_{j=1}^{n} \pi_j x_j \\ &\text{subject to } \textstyle\sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad && i = 1, \dots, m \qquad (1) \\ &x_j \in [0,1] && j = 1, \dots, n, \end{aligned}$$

where m denotes the number of capacity constraints and n represents the number of columns. While a_{ij} stands for the requested items by bidder j regarding resource i with capacity b_i, the term π_j denotes the willingness-to-pay for the total package. In an online auction the coefficients $(\pi_j, \boldsymbol{a_j})$ are revealed consecutively over time. The contribution of [1] is to compute the dual solution of a partial linear program and use it as a decision rule for future incoming bidders. The key idea is to set threshold prices $\boldsymbol{p} = (p_i)_{i=1,..,m}$ for each resource i equal to the dual prices of a linear program that is solved after a fraction of $\varepsilon \in (0,1)$ columns are revealed. Subsequent incoming bids $(\pi_j, \boldsymbol{a_j})$ will be compared to the current threshold prices $\boldsymbol{p}$. A bid will only be accepted if π_j exceeds $\boldsymbol{p^T a_j}$ and if no resource constraint is violated, meaning the problem remains feasible with $x_j = 1$.

OLA:	DLA:	(2)
maximize $\sum_{j=1}^{s} \pi_j x_j$	maximize $\sum_{j=1}^{l} \pi_j x_j$	
s.t. $\sum_{j=1}^{s} a_{ij} x_j \leq (1-\varepsilon)\frac{s}{n} b_i$	s.t. $\sum_{j=1}^{l} a_{ij} x_j \leq (1-h_l)\frac{l}{n} b_i$	$i = 1, \dots, m$
$x_j \in [0,1]$	$x_j \in [0,1]$	$j = 1, \dots, s$

[1] present two algorithms: the so called *One-Time Learning Algorithm* (OLA) and the *Dynamic Learning Algorithm* (DLA). The respective partial linear programs for retrieving the sets of dual prices are denoted above. The OLA learns a single threshold price vector at time $s = \varepsilon n$, applicable to all following bids. All incoming requests until s are only used to calculate the threshold prices and will always be rejected. Assuming that n is known and $(\pi_j, \boldsymbol{a_j})$ arrive in random order, [1] show that the OLA exhibits $(1 - 6\varepsilon)$-competitiveness against the ex-post optimum (OPT) under the right-hand side condition $B = \min_i b_i \geq \frac{6m\log(n/\varepsilon)}{\varepsilon^3}$.

The DLA updates dual prices in geometric intervals at $\varepsilon n, 2\varepsilon n, 4\varepsilon n, 8\varepsilon n, \dots$, in particular for each $l = \lceil 2^r \varepsilon n \rceil$ with the largest integer r such that $l < j$. The right-hand side of the partial linear programs is modified by a factor $h_l = \varepsilon\sqrt{\frac{n}{l}}$. The DLA is $1 - O(\varepsilon)$ competitive given a milder condition $B = \min_i b_i \geq \frac{m\log(n/\varepsilon)}{\varepsilon^2}$.

Note that both algorithms are distribution-free. However, [1] need ex-ante knowledge about the size of n in order to calculate the learning fractions of the respective algorithm. Furthermore, [1] assume that the bids $(\pi_j, \boldsymbol{a_j})$ arrive in random order. The latter assumption seems reasonable from a practical perspective, as the order of columns usually appears to be independent of the columns' content.

3 Experimental Analysis

3.1 Experimental Design

The main contribution of [1] is the introduction and (theoretical) analysis of these two algorithms for solving online allocation problems. However, they do not feature any specific implementation or application. This paper intends to examine the implied allocation mechanism from a practical perspective and to analyze sensitivities.

For this purpose, the algorithms are implemented and systematically tested in numerical experiments. In particular, all input parameters are varied in an orderly fashion to isolate important influencing factors on the algorithms' outcome and runtime. The treatment variables include the number of resources m, resource capacities b, initial learning fraction ε as well as number of bidders n. Moreover, building on the distribution-free property of the algorithms, we examine the robustness with respect to different stationary input distributions or distribution parameters regarding the willingness-to-pay π_j and the item requests a_{ij}.

Table 1. Parameters and Treatment Variables

	Description	*Values and Distributions*
Parameters	Simulations per Configuration	10
	Permutations per Simulation	100
	$\max(a_{ij})$	5
	a_{1j}, a_{2j}, a_{3j}	$N(0.4, 0.3), N(0.5, 0.6), N(0.6, 0.9)$
	$\pi_{1j}, \pi_{2j}, \pi_{3j}$	$N(100, 30), N(75, 20), N(50, 10)$
Treatment Variables	Resources m	$\{1; 2; \mathbf{3}; 5; 10\}$
	Capacity b	$\{50; 100; \mathbf{200}; 300; 400; 500; 600; 700; 800; 900; 1{,}000\}$
	Fraction ε	$\{0.001; \mathbf{0.01}; 0.02; 0.03; 0.04; 0.05; 0.075; 0.1; 0.125; 0.15; 0.2; 0.25\}$
	Bidders n	$\{200; 400; 600; 800; \mathbf{1{,}000}; 1{,}200; 1{,}400; 1{,}600; 1{,}800; 2{,}000; 5{,}000; 10{,}000\}$

The treatments are analyzed with respect to runtimes as well as objective function values, i.e. the revenue generated from the allocation in percent of the ex-post optimum. We define a baseline treatment which serves as a reference point for the subsequent ceteris paribus variation of treatment variables. The values and ranges for each parameter and variable are summarized in Table 1. The respective baseline configuration ($m = 3$, $b = 200$, $\varepsilon = 0.01$, $n = 1{,}000$) is highlighted in bold characters. All resources are assumed to have an identical capacity $b_i = b$. The parameter a_{ij} is required to be an integer value between zero and a pre-specified upper limit. In order to map distinct resource classes, the parameter a_{ij} follows a normal distribution with stepwise mean-variance-tuples. Similarly, the willingness-to-pay decrease over the different resources and are summed up to an aggregate π_j. Every bidder j is obligated to request at least one item.

Each treatment variable is considered in an isolated fashion. Under the ceteris paribus assumption, each input parameter is varied according to Table 1 using the baseline configuration as a starting point. Hence, our experimental design comprises 40 treatment combinations. For each configuration, 10 instances are created. Since [1] consider expected values over all permutations, each instance consists of 100 sub-instances, representing different permutations of the set of bidders.

Regarding the runtime evaluation, it should be noted that single permutations might be subject to noise distorting the measured runtimes. However, since averages are taken over a multitude of permutations, valid statements on general trends are ensured. Furthermore, we only intend to examine diverging runtime magnitudes and to identify common patterns. In order to explicitly exclude the possibility of invalid runtimes, we checked the results against larger problem instances. Essentially, we scaled the test configurations up by a factor of 100 (e.g. $b = 20{,}000$, $n = 100{,}000$ with 10 permutations and 5 simulations as the baseline treatment) and computed the respective ratios of runtimes. These ratios also took values of the up-scaling factor of 100, which we found to be consistent with our previous runtime analysis. Absolute runtimes should, however, always be handled with care according to [15].

In order to evaluate the objective function values and runtimes of both the OLA and the DLA, several simple benchmarks have been implemented as alternative measures. The most simplistic mechanism would be to accept any incoming request, as long as no constraint is violated. This quick and easy allocation principle is referred to as *Greedy Algorithm* in our context. In contrast, a so-called *Interval Learner* updates dual prices at constant 10%-intervals with respect to the ex-ante known number of incoming bids n. This benchmark is expected to always exhibit longer runtimes. The last benchmark, the *Willingness-To-Pay (WTP) Learner,* is based on the idea that it might be reasonable to update dual prices whenever the average over all bids per item changes by a certain magnitude, in our case by at least 5% in absolute terms. That is, if the pool of bidders appears to become more heterogeneous, different threshold prices might be deemed appropriate, calling for a re-calibration of dual prices. A fraction of 10% of the bids is used to calculate the initial dual prices.

The simulations were implemented in Python 3.6. Linear programs were solved using the interface to Gurobi 8.0. All simulations were performed on an Intel Core i7 7700K 4.20GHz quad-core machine with 32 GB of RAM.

3.2 Results

Starting with the baseline treatment, the well-known trade-off between approximation capabilities regarding the OPT and average runtimes already becomes visible. The numerical results indicate that the DLA excels in terms of the objective function (93.68% of the OPT). Evidently, the frequent update of dual prices incorporated into the DLA generates considerably better objective function values as opposed to the simpler OLA (69.42%), at least given the input parameters at hand. Furthermore, in spite of several updating steps of dual prices, the DLA certainly seems to be able to balance objective function output and runtime ($126.47ms$; OLA: $8.36ms$) reasonably. In contrast, while the *Greedy Algorithm* ($80.27\%; 1.53ms$) and the *WTP Learner* ($86.25\%; 97.31ms$) are executed faster, they cannot provide comparable approximations of the OPT due to their naivety. Likewise, the *Interval Learner* may come closer to the DLA vis-à-vis the objective (88.96%), but requires significantly more runtime ($405.35ms$). The OLA only proves to be competitive for large problem instances. Most notably, its approximation ratio increases to 95.16% in the up-scaled control scenario, potentially because the number of bids εn used for learning the dual prices is bigger in absolute terms. Therefore, the calibrated thresholds might be more valid than in the small-scale case.

Result 1. *The DLA provides a superior balance between objective function values and average runtimes. While several re-calibrations of dual thresholds enable good approximations of the ex-post optimum, the geometric updating intervals shift the majority of computational intensity toward small-scale optimization problems.*

Under the ceteris paribus condition, a variation of the total number of resources m, each having a capacity of 200 units, does not significantly affect the approximation performance of either the OLA or the DLA, as can be seen in Table 2. Employing a very little or a very large number of resources does not deteriorate the outcome substantially, potentially since each added resource inevitably comes along with new demand, as guaranteed by the simulation specifics. The performance of the *Greedy Algorithm*, however, constantly drops with an increasing number of resources. Table 2 also reports linear regression slope coefficients and adjusted R^2 figures for the objective function and the runtime. Regarding these coefficients, note that the dependent variables are denoted in percent and milliseconds (ms), respectively. Moreover, a significance level of 1% is chosen. The negative linear objective function sensitivity of the *Greedy Algorithm* may be ascribed to growing heterogeneity among bidders as a result of more resources and hence more scope of simulation. Since more deliberate decisions are necessary for a diverse pool of bidders, the naïve *Greedy Algorithm* cannot sustain its approximation ratio. The runtimes for almost all mechanisms increase significantly and in a linear fashion once new resources are added.

Because an increase in the number of resources goes hand in hand with newly generated demand, resource scarcity or abundance can better be reflected by changing the available capacities b. As displayed in Table 3, the *Greedy Algorithm*, the DLA, and the OLA ultimately converge to the OPT in terms of the objective function value when resources are excessively available. In particular, a 100%-approximation of the *Greedy Algorithm* indicates that all incoming bidders can be served. In this case, shadow prices may be close to zero and the other mechanisms exhibit a gap to the OPT mainly due to the initial calibration period, where all requests are rejected. When resources are scarce, however, the algorithms show significant discrepancies. As items need to be assigned with consideration, simple allocation mechanisms, as implied by the OLA or the *Greedy Algorithm*, produce below average results. The OLA, however, exhibits the steepest linear growth with increasing resource capacity. In contrast, the DLA already performs very well for limited availabilities. It also exhibits a slightly quadratic relationship, i.e. performing a quadratic regression increases the $\bar{R}^2$ from 86.45% in the linear case to 96.49% with the quadratic regression coefficient being statistically significant at the 1% level. That is, the DLA dominates all other benchmarks as long as resources are exposed to scarcity to some extent. It also exceeds the *WTP Learner* and the *Interval Learner*, where the first 10% of the bidders will always be rejected. Changing b does not significantly affect the average runtimes aside from the OLA and the *Greedy Algorithm*. That is, if more resource capacities are available, more bidders can be served, leading to a consistent upward trend in runtime, albeit on a small level. For the other mechanisms, this effect does not become visible, as it is only a tiny proportion of the total runtime.

The fraction ε represents an interesting lever for training-based algorithms, determining the number of bids initially required for calibrating the thresholds. As can be seen in Table 4, the three benchmarks exhibit zero sensitivity, as they do not make use of this parameter. In terms of the objective function value, the DLA shows the greatest dependence with respect to ε. Generally speaking, the smaller this fraction is chosen, the more learning instances are executed by the algorithm, enabling a better approximation of the OPT. For large ε, the OLA produces better results than the dynamic mechanism, indicating that the DLA is considerably restricted by the modifying right-hand side factor postulated by [1], artificially increasing the dual prices at each re-calibration for too large fractions. At the same time, unlike the DLA exhibiting a linearly decreasing behavior, the OLA produces its best results for a medium $\varepsilon = 0.05$. Since it only learns dual prices at a single time, a very small fraction of the sample will not be representative enough and thus result in a poor or incalculable performance. Therefore, in order to produce a proper outcome, a certain minimum share of bids needs to constitute the training set. The OLA thus exhibits a significant quadratic relationship with an $\bar{R}^2$ of 85.69%. Furthermore, the tests reveal linearly increasing OLA runtimes, since the single optimization problems will encompass more elements with increasing ε. As evident from the numerical results, the shape of the DLA runtime function is rather serrated. Holding the total number of required optimization steps constant, the runtime would increase with growing ε for the same reasons as the OLA. However, once ε exceeds some threshold, one former optimization step is not feasible anymore, thereby reducing the total number of dual price updates and significantly decreasing the average runtime.

Table 2. Numerical Results for Treatment Variable m

Resources m	*1*	*2*	*3*	*5*	*10*	β	$\bar{R}^2$
DLA	91.06%	93.64%	93.77%	92.49%	88.69%	-	-
	118.20ms	165.96ms	209.76ms	291.21ms	503.05ms	42.43	99.96%
OLA	70.21%	71.20%	69.24%	69.27%	65.08%	-	-
	9.49ms	12.78ms	14.45ms	18.05ms	27.70ms	1.95	99.31%
Greedy	87.65%	86.22%	79.78%	72.96%	62.55%	-2.85	95.19%
Algorithm	1.07ms	2.19ms	2.68ms	3.35ms	5.40ms	0.444	95.57%
Interval	84.24%	88.39%	89.27%	89.50%	87.89%	-	-
Learner	361.09ms	517.65ms	662.04ms	942.33ms	1675.44ms	145.33	99.99%
WTP	83.33%	86.79%	86.09%	85.20%	81.25%	-	-
Learner	42.10ms	146.69ms	160.14ms	221.53ms	439.07ms	40.83	96.52%

Table 3. Numerical Results for Treatment Variable b

Capacities b	*50*	*100*	*200*	*300*	*400*	*500*	*600*	*700*	*800*	*900*	*1,000*	β	$\bar{R}^2$
DLA	89.43%	91.78%	93.73%	94.70%	95.83%	96.52%	97.36%	97.78%	97.77%	98.59%	98.92%	0.00868	86.45%
	225.52ms	221.24ms	222.40ms	221.65ms	222.79ms	223.30ms	223.83ms	224.49ms	227.53ms	226.07ms	223.96ms	-	-
OLA	67.56%	63.57%	70.42%	74.67%	79.92%	83.77%	86.90%	90.32%	92.21%	94.93%	96.96%	0.0351	96.78%
	14.97ms	14.99ms	16.18ms	16.68ms	17.52ms	17.93ms	18.70ms	19.50ms	19.86ms	20.01ms	20.19ms	0.00595	96.72%
Greedy	73.40%	77.87%	79.73%	79.77%	83.10%	83.64%	87.27%	92.26%	95.79%	100.00%	100.00%	0.0278	96.07%
Algorithm	2.54ms	3.08ms	3.59ms	4.31ms	5.00ms	4.84ms	4.97ms	4.73ms	5.54ns	5.22ms	5.46ms	0.00274	77.88%
Interval	86.93%	88.03%	89.04%	89.53%	89.39%	89.47%	89.57%	89.70%	89.72%	89.89%	89.99%	0.00230	61.23%
Learner	710.78ms	696.32ms	693.92ms	689.87ms	691.53ms	689.57ms	690.73ms	693.39ms	701.08ms	690.55ms	686.45ms	-	-
WTP	80.73%	83.43%	86.34%	87.80%	88.23%	88.88%	89.25%	89.61%	89.59%	89.74%	89.99%	0.00784	69.40%
Learner	176.46ms	173.45ms	174.65ms	170.45ms	175.35ms	169.06ms	172.63ms	178.59ms	176.45ms	170.80ms	172.75ms	-	-

Table 4. Numerical Results for Treatment Variable ε

Fraction ε	*0.001*	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*	*0.075*	*0.1*	*0.125*	*0.15*	*0.2*	*0.25*	β	$\bar{R}^2$
DLA	94.54%	93.67%	92.48%	90.06%	88.53%	86.32%	80.84%	75.88%	70.36%	65.16%	56.39%	47.25%	-196.66	99.82%
	109.82ms	125.89ms	123.25ms	176.91ms	120.06ms	146.36ms	108.35ms	140.16ms	85.01ms	99.91ms	129.10ms	72.08ms	-	-
OLA	76.12%	70.19%	76.47%	78.81%	80.47%	81.49%	80.87%	79.79%	76.26%	73.28%	65.87%	58.22%	-62.73	45.92%
	8.35ms	8.26ms	9.12ms	9.81ms	10.63ms	11.46ms	13.38ms	15.33ms	17.25ms	19.23ms	23.20ms	27.09ms	77.39	99.87%
Greedy	80.67%	79.71%	80.26%	79.61%	79.88%	80.13%	79.91%	80.85%	79.14%	80.08%	80.02%	80.11%	-	-
Algorithm	1.54ms	1.54ms	1.53ms	1.54ms	1.54ms	1.51ms	1.53ms	1.56ms	1.54ms	1.54ms	1.54ms	1.52ms	-	-
Interval	88.97%	88.98%	89.15%	88.76%	88.95%	89.07%	89.06%	89.10%	89.10%	88.82%	88.91%	88.83%	-	-
Learner	404.87ms	403.81ms	403.80%	404.97ms	403.68ms	404.70ms	403.79ms	404.57ms	403.14ms	404.58ms	404.65ms	404.70ms	-	-
WTP	86.28%	86.39%	86.36%	86.11%	86.14%	86.45%	86.21%	86.48%	86.16%	86.14%	86.38%	86.32%	-	-
Learner	101.45ms	100.35ms	98.78ms	97.56ms	99.01ms	98.01ms	99.38ms	97.96ms	97.67ms	99.96ms	99.44ms	97.07ms	-	-

Table 5. Numerical Results for Treatment Variable n

Bidders n	*200*	*400*	*600*	*800*	*1,000*	*1,200*	*1,400*	*1,600*	*1,800*	*2,000*	*5,000*	*10,000*	β	$\bar{R}^2$
DLA	98.36%	95.30%	93.88%	93.64%	93.64%	93.56%	93.45%	93.18%	93.48%	93.42%	92.98%	92.54%	-	-
	29.96ms	54.14ms	78.36ms	102.48ms	126.14ms	150.63ms	174.37ms	198.78ms	222.92ms	247.69ms	607.05ms	1208.25ms	0.120	100.00%
OLA	86.91%	77.39%	71.91%	70.33%	69.62%	68.74%	69.36%	69.32%	69.11%	69.22%	69.88%	69.83%	-	-
	2.89ms	4.18ms	5.58ms	6.94ms	8.24ms	9.56ms	10.91ms	12.19ms	13.48ms	14.83ms	33.95ms	65.72ms	0.00640	99.99%
Greedy	100.00%	84.44%	81.01%	79.57%	79.84%	80.24%	81.25%	80.10%	78.94%	78.33%	71.80%	68.25%	-0.00194	45.70%
Algorithm	0.47ms	0.82ms	1.09ms	1.33ms	1.55ms	1.74ms	1.95ms	2.14ms	2.32ms	2.50ms	5.03ms	9.25ms	0.00087	99.64%
Interval	89.83%	89.23%	89.14%	89.11%	89.10%	88.75%	88.83%	88.35%	88.59%	88.44%	87.94%	87.15%	-0.00023	79.86%
Learner	84.96ms	166.57ms	246.56ms	325.99ms	404.26ms	484.82ms	562.31ms	641.81ms	720.15ms	800.69ms	1987.33ms	3973.44ms	0.396	100.00%
WTP	89.73%	87.88%	87.09%	86.59%	86.48%	85.83%	85.94%	85.23%	85.53%	85.24%	84.52%	84.32%	-	-
Learner	29.51ms	49.00ms	65.75ms	82.25ms	97.93ms	114.43ms	128.86ms	150.19ms	163.79ms	182.18ms	419.64ms	828.72ms	0.0811	99.99%

Changing the number of stationary bidders n competing for the fixed resource capacities, the DLA once again proves to be robust. This can be seen in Table 5. While it becomes increasingly difficult to select the ex-post optimal requests from a larger set of bidders, the drop in performance is not as significant as with other benchmarks. The *Greedy Algorithm*, for instance, cannot maintain its approximation ratio. The *WTP Learner* and the *Interval Learner* remain robust to a certain extent, yet never reaching the performance of the DLA. The OLA, in turn, is more sensitive towards changes in n. With a growing number of requests, a single learning step is not sufficient due to the multitude of bidders arriving subsequent to the calibration. The average runtimes of all algorithms naturally increase with problem size. The functional relationship concerning the runtimes is linear for all mechanisms.

We validated the observations of these ceteris paribus analyses with a multivariate parameter grid and were not able to detect major deviations. The results above might, however, also be contingent upon the distribution assumptions. In general, there are six levers to be tested with respect to the stationary input processes: mean and standard deviation of the willingness-to-pay π_j and of the requests a_{ij}, given a normality assumption, as well as the underlying distributions of the simulated parameters themselves. As pointed out by [1], the DLA stands out due to its distribution-free property and thus robustness to all tuning parameters. Conversely, the *Greedy Algorithm* benefits from homogeneous pools of bidders, exhibiting major improvements for decreasing standard deviations of π_j. It is also most sensitive to changes of the underlying distributions of π_j and a_{ij}. The normality assumption appears to be most suitable for achieving little runtimes. If the distribution parameters fluctuate within a simulation run, the DLAs performance deteriorates, yet is still competitive to the benchmarks.

Result 2. *The DLA proves to be robust regarding changes of most input parameters. Its capability to produce near-optimal objective function values remains widely unaffected by changes in the number of resources, available resource capacities, number of bidders, or distribution assumptions. The DLA thus seems to be applicable to any kind of online auction configuration with a revenue maximization objective.*

Result 3. *The average runtimes of the DLA depend linearly on the number of resources and bidders. They are stable with respect to resource capacities. These statements also apply to the OLA and more naïve benchmarks.*

Result 4. *The initial learning fraction ε represents the most critical sensitivity for the DLA. Small fractions are necessary in terms of the objective function value, but come at the cost of more computing time. The indispensable refusal of the first εn bidders is a key drawback with respect to the practical applicability of the DLA.*

Given these experimental results, several interesting inferences can be drawn with respect to the practical applicability of the DLA. In terms of the approximation of the OPT, the findings above indicate that large capacities b, small fractions ε, reasonable numbers of resources m, and few bidders n favor the DLA. The positive discrepancy towards other mechanisms, however, seems to increase for smaller b and larger n, i.e. conditions of resource scarcity and excess demand. Customer heterogeneity, implemented through larger standard deviations for π_j, also contributes to the superiority of this mechanism. In order to validate these statements, a modified baseline treatment is defined with the main purpose of making the DLA the most dominant algorithm. Here, excess demand ($n = 2{,}000$) together with a little learning fraction ($\varepsilon = 0.001$) and more heterogeneous bidders ($\sigma_{\pi_{ij}} = (60{,}40{,}20)\ \forall j$) are selected, while all other input figures remain identical to the initial baseline treatment. In this configuration, the DLA distinctly outperforms all other benchmarks with a 94.60%-approximation of the OPT. The *Interval Learner* follows with a ratio of 89.10%, but with more than thrice the runtime of the DLA. While the *WTP Learner* reaches 86.42%, the *Greedy Algorithm* and the OLA only produce 64.56% and 64.51%, respectively. Hence, the DLA excels in terms of OPT-approximation when the allocation is complex, e.g. if resources are scarce and bidders heterogeneous, and thus complies well with the nature of the generic online tickets sales problem. The average runtimes do not differ from our initial findings.

Result 5. *The DLA is especially suited for non-trivial online auction problems, in particular for scenarios with resource scarcity and excess demand by heterogeneous bidders. Its degree of optimality is highly contingent upon the choice of the initial learning fraction ε. Subject to these limitations, the exemplary case of online ticket sales appears to be a viable and expedient area of application.*

The key drawback of the DLA seems to be the unconditional refusal of the first εn bidders, as this fraction is required to learn the first set of threshold prices. Especially for large ε, it might be advantageous to define initial threshold prices to not lose revenue from the learning fraction. Moreover, the right-hand side modifier introduced by [1] artificially increases the dual prices, leading to overly restrictive thresholds for very large ε. Early-arriving bidders might also feel discriminated due to their arbitrary rejection. Since limited computing power or other conceivable reasons might prohibit the employment of sufficiently small learning fractions, this downside needs to be addressed in order to improve the applicability of the DLA for practical use cases, for instance for online auction-based ticket sales.

As explained by [1], the DLA is explicitly designed as a distribution-free mechanism. This seems a particularly useful property if no knowledge is available about the incoming bidders. If the allocation is repeated regularly, however, it might be possible to infer estimators for distribution parameters from historical data. In this case, one should define some initial threshold prices based on the given information about the stationary input processes in order to mitigate the problems associated with a big ε. Therefore, a deterministic linear program (LP), in particular replacing stochastic variables with their expected values, might be utilized for estimating meaningful thresholds. Dual prices can be accessed after solving the LP with expected willingness-to-pay and item requests. These shadow prices should be good references points for initial thresholds as they mirror expected values. Because all bidders are homogeneous in expectation, the LP can be solved by accepting all requests as long as no capacity restrictions are violated, just like the *Greedy Algorithm*. Since two resources will still be available when the third one is exhausted due to our assumption of stepwise demands, only one dual price will take a positive value in our setting.

If the allocation problem has to be solved repeatedly with similar input features, it might also be reasonable to use an average over past dual prices as initial thresholds. In particular, when input distributions are known, several simulations can be executed prior to the actual allocation task. For each simulation run, the first set of dual prices after εn bidders is stored. The average over this set of shadow prices, in our case over 10 simulations, can serve as initial thresholds for the actual allocation.

Alternatively, more naïve initial threshold prices are the expected values of the known input distributions for the willingness-to-pay for each resource. All bids, exceeding these prices will be accepted as long as the problem remains feasible.

Table 6. OPT-Approximations with Initial Thresholds

ε	*Baseline Treatment*	*LP Thresholds*	*Simulated Thresholds*	*Expected Values of Distributions*
0.001	94.54%	94.41%	94.74%	94.44%
0.01	93.67%	94.23%	94.84%	94.53%
0.03	90.06%	92.86%	91.66%	93.49%
0.05	86.32%	91.69%	91.10%	92.31%
0.1	75.88%	87.33%	85.75%	89.44%
0.15	65.16%	82.63%	77.95%	86.10%
0.25	47.25%	74.93%	63.51%	80.85%

The results of these different approaches for different ε are presented in Table 6. While the conventional DLA begins to rapidly deteriorate for $\varepsilon \geq 0.05$, initial thresholds keep the performance on a higher level. In particular, using thresholds does not seem to be disadvantageous for any ε. Simulated dual prices appear to be too instance-specific, making it difficult to rely on a limited set of past thresholds for future allocations. Retrieving initial thresholds from solving a LP produces better approximations of the OPT. This especially applies to scenarios where large fractions ε are chosen. Simply employing expected values of the distributions, however, further improves the approximation capabilities and should even be preferred to the more sophisticated alternatives. Again, it should be emphasized that this extension does not come along with any essential drawbacks. For small ε only few bidders are affected, whereas initial thresholds enable major gains for large fractions.

The DLA together with well-defined initial thresholds can therefore be deemed appropriate for complex allocation tasks. It represents a viable alternative to online ticket sales or other multi-item B2C businesses, achieving superior approximations of the OPT while maintaining satisfactory runtimes at the same time.

Result 6. *Defining initial threshold prices by using available information on the input distributions mitigates the drawbacks of the DLA associated with very large learning fractions ε. Matching the thresholds with the expected willingness-to-pay until the first calibration of dual prices already aids in alleviating revenue loss. It can also improve on the perceived fairness in handling the bidders and thus enhances the practical applicability of the DLA substantially.*

4 Conclusion

Online auctions represent a promising alternative to conventional posted price mechanisms, potentially enabling a more effective exploitation of customer willingness-to-pay. Based on the notion of ticket sales, we conducted an experimental study on two seminal algorithms proposed by [1] and put special emphasis on the practicability of their underlying primal-dual framework. Both the OLA and the DLA were implemented for the purpose of simulation-based experimental testing, along with some intuitive benchmarks, ranging from a quick-and-easy *Greedy Algorithm* to a computationally intensive *Interval Learner*.

There is a fundamental trade-off between the capability to approximate the ex-post optimal revenue and the average computational runtime. We ran extensive numerical experiments to discover dependencies and sensitivities of these opposing objectives. Through precisely defined re-calibrations of dual threshold prices, the DLA is able to approximate the OPT very well against a stationary process of bids. At the same time, it maintains reasonable runtimes, since dual updates occur more frequently at early stages of the allocation process. Our experiments illustrate that the DLA reacts robust to changes in many input parameters and proves to be extraordinarily dominant in situations of resource scarcity and excess demand. In this case, the decision rules implied by the DLA enable deliberate allocations through accurately determined thresholds.

Addressing the problem of generally rejecting bids in the first learning phase, we drafted several extensions, making use of known distribution information and aiming at the definition of some initial thresholds that could be employed until the first ordinary calibration. While retrieving a set of dual prices from a deterministic linear model seems an elegant solution, simply using the expected values of the known input distributions as initial thresholds proves to be an easy and well-performing alternative. Initial threshold prices permit major enhancements with respect to the real-world applicability of the DLA. Together with its robustness regarding other treatment variables, the DLA can thus be viewed as an expedient alternative when it comes to online ticket sales or other revenue-maximizing multi-item allocation tasks.

The basic idea underlying this research paper can be extended in various ways. An empirical validation with real-world data would, for instance, be desirable. Especially the context of online ticket sales seems to be a suitable use case. Furthermore, the algorithms have only been benchmarked with rather straightforward mechanisms. A more extensive experimental study with alternative sophisticated algorithms might provide useful insights into the applicability of different online auction mechanisms. Including risk considerations or non-monetary objectives would only be two possible extensions referring to the DLA.

References

1. Agrawal, S., Wang, Z., Ye, Y.: A Dynamic Near-Optimal Algorithm for Online Linear Programming. Operations Research 62 (4), 876–890 (2014)
2. Babaioff, M., Immorlica, N., Kempe, D., Kleinberg, R.: Online Auctions and Generalized Secretary Problems. ACM SIGecom Exchanges 7 (2), 1–11 (2008)
3. Buchbinder N., Jain K., Naor J.: Online Primal-Dual Algorithms for Maximizing Ad-Auctions Revenue. In: Arge, L., Hoffmann, M., Welzl, E. (eds.) Algorithms – ESA 2007. LNCS, vol. 4698, pp. 253–264. Springer, Berlin, Heidelberg (2007)
4. Buchbinder N., Naor J.: Improved Bounds for Online Routing and Packing via a Primal-Dual Approach. In: 2006 47th Annual IEEE Symposium on Foundations of Computer Science, pp. 293–304. FOCS '06, Berkeley (2006)
5. Buchbinder N., Naor J.: Online Primal-Dual Algorithms for Covering and Packing. Mathematics of Operations Research 34 (2), 270–286 (2009)
6. Cai, Y., Devanur, N.R., Weinberg, S.M.: A Duality Based Unified Approach to Bayesian Mechanism Design. ACM SIGecom Exchanges 15 (1), 71–77 (2016)
7. Devanur, N.R.: Online Algorithms with Stochastic Input. ACM SIGecom Exchanges 10 (2), 40–49 (2011)
8. Devanur, N.R., Hayes T.P.: The Adwords Problem: Online Keyword Matching with Budgeted Bidders under Random Permutations. In: Proceedings of the 2009 ACM Conference Electronic Commerce, pp. 71–78. EC '09, Stanford (2009)
9. Dütting, P., Feldman, M., Kesselheim, T., Lucier, B.: Prophet Inequalities Made Easy: Stochastic Optimization by Pricing Non-Stochastic Inputs. In: 2017 IEEE 58th Annual Symposium on Foundations of Computer Science, pp. 540–551. FOCS '17, Berkeley (2017)
10. Feldman, J., Henzinger, M., Korula, N., Mirrokni, V., Stein, C.: Online Stochastic Packing Applied to Display Ad Allocation. In: de Berg, M., Meyer, U. (eds.) ESA 2010, Part I. LNCS, vol. 6346, pp. 182–194. Springer, Berlin, Heidelberg (2010)
11. Feldman, J., Korula, N., Mirrokni, V., Muthukrishnan, S., Pál, M..: Online Ad Assignment with Free Disposal. In: Leonardi, S. (ed.) WINE 2009. LNCS, vol. 5929, pp. 374–385. Springer, Berlin, Heidelberg (2009)
12. Gönsch, J., Klein, R., Neugebauer, M., Steinhardt, C.: Dynamic Pricing with Strategic Customers. Journal of Business Economics 83 (5), 505–549 (2013)
13. Hajiaghayi, M.T., Kleinberg, R., Sandholm, T.: Automated Online Mechanism Design and Prophet Inequalities. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, pp. 58–65. AAAI '07, Vancouver (2007)
14. Kleinberg, R.: A Multiple-Choice Secretary Algorithm with Applications to Online Auctions. In: Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 630–631. SODA '05, Vancouver (2005)
15. Kriegel, H.P., Schubert, E., Zimek, A.: The (Black) Art of Runtime Evaluation: Are We Comparing Algorithms or Implementations? Knowl Inf Syst 52 (2), 341–378 (2017)
16. Strauss, A.K., Klein, R., Steinhardt, C.: A Review of Choice-Based Revenue Management: Theory and Methods. European Journal of Operational Research 271 (2), 375–387 (2018)
17. Zhong, L., Huang, Q., Wu, F., Chen, G.: TRADE: A Truthful Online Combinatorial Auction for Spectrum Allocation in Cognitive Radio Networks. Wireless Communications and Mobile Computing 15 (9), 1320–1330 (2015)

Design Features of Non-Financial Reward Programs for Online Reviews: Evaluation based on Google Maps Data

Alexander Kupfer[1], Luciano van Essen[2] and Steffen Zimmermann[3]

[1]University of Innsbruck, Department of Banking and Finance, Innsbruck, Austria
(alexander.kupfer)@uibk.ac.at
[2,3]University of Innsbruck, Department of Information Systems, Production and Logistics Management, Innsbruck, Austria
(luciano.van-essen,steffen.zimmermann)@uibk.ac.at

Abstract. Many online review platforms such as Google Maps or TripAdvisor have implemented a reward program where reviewers receive points for posting online reviews. Based on the Self-Determination Theory, we hypothesize that introducing multiple reward levels and incentives for detailing online reviews have a positive effect on review quantity and effort. We test our hypotheses for Google's reward program 'Local Guides' which introduced new design features in June 2017. At this date, more badges (i.e., reward levels) were available and extra points for textual reviews (i.e., incentives for detailing online reviews) were provided. Based on a dataset of 43,988 online reviews on 40 sights across Europe, we find that multiple reward levels and incentives for detailing reviews significantly increase the number of reviews but counterintuitively decrease the effort invested by reviewers. Consequently, the introduction of extra points for textual reviews has not yield the intended effect on review effort.

Keywords: Online Review Platform, Reward Program, Self-Determination Theory, Review Effort, Review Quantity.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Topic Embeddings – A New Approach to Classify Very Short Documents Based on Predefined Topics

Lasse Lommel[1], Meike Riebeling[1], Burkhardt Funk[1] and Christian Junginger[2]

[1] Leuphana University Lüneburg, Institute of Information Systems, Lüneburg, Germany
{lasse.lommel,meike.b.riebeling}@stud.leuphana.de,
burkhardt.funk@leuphana.de
[2] Otto GmbH & Co KG, Data Science, Hamburg, Germany
christian.junginger@otto.de

Abstract. Traditional unsupervised topic modeling approaches like Latent Dirichlet Allocation (LDA) lack the ability to classify documents into a predefined set of topics. On the other hand, supervised methods require significant amounts of labeled data to perform well on such tasks. We develop a new unsupervised method based on word embeddings to classify documents into predefined topics. We evaluate the predictive performance of this novel approach and compare it to seeded LDA. We use a real-world dataset from online advertising, which is comprised of markedly short documents. Our results indicate the two methods may complement one another well, leading to remarkable sensitivity and precision scores of ensemble learners trained thereupon.

Keywords: topic modeling, word embeddings, LDA, seeded LDA

1 Introduction

With the increasing amount of textual data, the interest in text analysis has grown significantly in recent years and methods like topic modeling have become an integral part of research in computer science and information systems [1-3]. Topic modeling aims at finding topics in a collection of text documents.

One popular unsupervised method in the field is Latent Dirichlet Allocation (LDA), which is a probabilistic model that infers a document's topic based on the distribution of words therein, building upon the underlying assumption that topics generate words, which in turn generate documents. It was developed from probabilistic latent semantic analysis (PLSA) by Blei et al. [4] and has since been one of the most widely adopted approaches to topic modeling, which is mirrored by its readily-available implementations in many programming languages and environments.

However, in some scenarios, one may — possibly due to domain knowledge — already know about a specific set of topics that is present within the corpus or that one would like to classify documents into for other reasons. Under such circumstances, unsupervised methods may fail to identify particularly those envisioned topics in a corpus of documents. Standard LDA, for example, takes input on only the *number* of topics to

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

discover but it does not necessarily find topics similar to those that one aims to classify into. Instead, the topics identified with the means of standard unsupervised methods would have to be further investigated by a human in order to identify the discovered topics' meaning, beyond them constituting a collection of associated documents, and possibly map them to a predefined topic scheme.

Supervised methods do not suffer from this problem since the labels establish the set of topics that can be discovered. Approaches of supervised learning suitable for topic modeling include the naive Bayes algorithm, support vector machines or decision tree models, which require attaching topic labels to words [5]. But also derivatives of LDA, namely supervised or discriminative LDA approaches (sLDA, DiscLDA) [6], [7], have been developed. However, all supervised models come at the cost of labeled data, which may be substantial, if not prohibitive in some cases. Moreover, in case the set of topics that one would like to categorize into changes, again significant amounts of newly generated labels or re-labeling of the data are required.

A solution to solve the dilemma of combining unsupervised methods with a predefined set of topics may lie in seeded LDA, a recent extension to the traditional form of LDA [8].

We develop an alternative method that employs recent developments in word vector representations. Our proposed method matches current undertakings in the field, as, for instance, Schwaiger et al. [9] and Murawski and Bick [10] have dealt with similar problems of classifying short documents into predefined sets. Schwaiger et al. [9] analyze social media posts and develop a classification tool, utilizing a dictionary-based approach combined with a multinomial Naive Bayes algorithm. In a similar endeavor, Murawski and Bick [10] employ LDA to sort job advertisements of data professionals into job sub-categories within specific job titles. However, their purely unsupervised approach yielded results that had to then be manually linked to the given category framework [10]. Further examples of topic modeling implementations are presented by Müller and Brocke [11] or Geva et al. [12]. In both papers, the authors use LDA to retrieve topics from a collection of documents. Müller and Brocke [11] apply LDA to further develop categories for an app store by analyzing the short descriptions provided for each application. Following a similar procedure, Geva et al. [12] analyze tweets to find patterns in the posting behavior of users.

Extended approaches to retrieve topic models are presented by few authors. Xu et al. [13] propose a combination of LDA and Clustering methods as a novel method to incorporate text associations. The approach is twofold. To find topic associations, a clustering algorithm based on the pairwise proximity of topics is applied. These associations are then linearly combined with LDA to enrich topic retrieval [13]. A very different direction for enriching topic modeling research is presented by Eickhoff and Wieneke [14]. Emphasizing the need to make sense of topics retrieved by topic modeling algorithms, they propose a mixed methods approach combining qualitative coding and quantitative clustering to decontextualize and evaluate topic models [14]. These examples underline the current focus on vanilla LDA, indicating a need for the development of specialized methods in topic modeling [1], [2].

We contribute to topic modeling research a novel alternative to (seeded) LDA, departing from the proven paths of generative probabilistic models. Our proposed method, which we present in section 2, is easy to set up and allows high performance classification of (very) short documents into predefined topic schemes and does not require labeled training data. In section 3, we verify our proposed approach on a dataset of 380,000 URL-based short documents that we obtain from an online retailer, where the classification of such documents into a predefined taxonomy allows for a better understanding of the websites on which online advertisement is served. Section 4 discusses the results, which suggest the proposed method may constitute a valuable addition to the existing topic modeling tool set. Section 5 concludes.

2 Methods

Section 2.1 provides an overview of related work from the field of (seeded) LDA. Subsequently, we present our novel approach to topic modeling which relies on word vector embeddings (section 2.2). We refer to this approach as topic embeddings.

2.1 Seeded LDA

Seeded Latent Dirichlet Allocation is a powerful probabilistic model in the field of unsupervised topic and text mining algorithms. It is an extension of the standard LDA model developed by Blei, Ng and Jordan [4]. In contrast to vector and matrix based text mining approaches such as latent semantic indexing or clustering, LDA algorithms not only reduce dimensionality, but also account for documents being composed by a mixture of topics [4], [15], [16]. The method is therefore suited for organizing large and unstructured text documents by extracting representative topic compositions [17]. Successful implementations are documented in a broad range of domains. In the field of information system research, examples include the analysis of social media to develop a tool categorizing social media activities or the analysis of computer science job advertisements, which we described in section 1 [9], [10], [15].

Jagarlamudi, Daumé and Udupa [8] developed the seeded LDA model for settings where true labels are not available but prior knowledge of topics inside the corpus can be used to steer the resulting topic assignments towards predefined topics. The general idea of seeded LDA is to feed the model with information about the nature of the topics to look for by integrating seed words. Setting such seeds in the LDA algorithm aims at guiding the extraction of topics by skewing the distributions on the document level as well as on the word level (document-topic distribution, topic-word distribution).

On the topic-word level, seed words enhance the probability of a topic to generate words which are associated with the seed words. The model is then extended by another topic-word distribution for each seed topic. Hence, a new layer is added, where each topic is assumed to be a mixture of its regular topic distribution and the related seed topic distribution. In addition to the topic-word level, seeded LDA guides the probability distributions in the document-topic layer. For each set of seeds, a distribution over the regular topics is built, which acts as a prior for drawing the document-topic distribution. Drawing a topic for each document is enhanced by an extra layer, where the first step is sampling a set of seeds and then using the corresponding distribution over topics as prior information to select the document topic mixtures [8].

In other words, seeded LDA is an unsupervised approach to direct topic mining towards a given set of categories. That is to say, it is a method which incorporates knowledge about desired labels in an unsupervised learning algorithm [8], [18]. However, it relies on a sufficiently large number of words per document to properly find topic models within a document collection. The algorithm is based on the probability of words to co-occur. As a result, the success of seeded LDA depends on the composition of the document collection and the quality of seed words, which not only have to differentiate a given topic from other topics but also must frequently occur in the corpus. However, for short documents or underrepresented topics, seeded LDA has its limitations.

2.2 Topic Embeddings

Word embeddings are high-dimensional representations of words in vector space. The words are translated into vectors of real numbers in such a way that semantically similar words result in similar vectors, where the translation is based upon evaluating how words co-occur with other words. Even though the basic idea of representing semantic similarities of words in vectors dates back to the 1960s, when the first attempts were made to use feature representations to capture such similarities between words [19], the interest in word embeddings was boosted recently when Mikolov et al. [20] introduced a neural networks-based model called word2vec in 2013, which permitted much faster training than previous models.

As words are translated into vectors, it becomes possible to determine numeric distances between words. Thus, one can calculate the distances between the words of a document and another term. This other term (or a collection of terms for that matter) could well constitute a topic, which leads to the core of our approach. It assigns each document to the closest topic in a given set of topics, where closeness is measured in the word embeddings' vector space. Thus, we dub our novel classification method *topic embeddings*.

However, not all words in a document are of equal importance. Stop words like "the" and other recurrent words add no value to determine a document's topic as they are neither unique nor characteristic to documents of any specific topic. To filter for the characteristic words that help to measure the distance between documents and topics, we use the *term frequency inverse document frequency (tf-idf)* metric. Tf-idf assigns weights to words within a corpus' documents and balances the relative frequency of a word within a document against the word's relative frequency in the whole corpus. High scores stand for characteristic words of a document which occur frequently (that is at least once) in the given document but relatively seldom in the corpus. To efficiently use tf-idf scores in our setting, we first compute a (sparse) document-term matrix that contains the tf-idf scores and normalize it row-wise so that the tf-idf scores per document sum up to 1, which makes them a suitable mean of weighting.

To assign a given document to a topic from a predefined set, we combine word embedding distances with tf-idf scores as follows: We measure the distances between all words in a given document and the terms describing a topic and combine these into a weighted sum, where the weights are contributed by the document words' normalized tf-idf scores.[1] We repeat this for all topics and assign the document to the closest topic:

Define `ClosestTopic`($\mathbf{x}$,$\mathbb{T}$,$\mathbb{C}$,F):
Require: $\mathbf{x}$, the input document with documents words $x^{(i)}$
Require: $\mathbb{T}$, the set of topics, consisting of topics $\boldsymbol{t}^{(j)}$ and topic words $t^{(j,k)}$
Require: $\mathbb{C}$, the corpus, comprised of a set of all documents
Require: *F(word, word)*, a distance function from a word embedding space
Calculate tf-idf document-word matrix $\mathbf{W}$ from $\mathbb{C}$.
Normalize $\mathbf{W}$ row-wise (per document) and extract $\mathbf{w}$ with weights $w^{(i)}$ for each document word $x^{(i)}$.
for j from 1 to $|\mathbb{T}|$ **do**

$$distance_j = \frac{\sum_{i=1}^{len(\boldsymbol{x})} \sum_{k=1}^{len(\boldsymbol{t})} w^{(i)} F(x^{(i)}, t^{(j,k)})}{len(\boldsymbol{t})}$$

$y \leftarrow argmin(distance_1, distance_2, \ldots, distance_{|\mathbb{T}|})$
Return topic $\boldsymbol{t}^{(y)}$

Words that are not in the vocabulary of the employed word embeddings model cannot be used, even though they may have high tf-idf scores. Documents that do not contain a single word from the word embedding model's vocabulary can thus not be classified at all.

Our approach shares elements with the work of De Boom et al. [21], who devise vector representations of fixed-length short documents with the help of idf-scores and word embeddings. Other approaches to obtain text representations include deep learning and skip-gram based methods. Multilayer perceptron and convolutional neural networks can be used to arrive at sentence representations as De Boom et al. [21] note. However, as De Boom et al. [21] point out, these methods necessitate inputs of the same length or aggregation operations. Other techniques to measure semantic sentence relationships by vector representations include skip-gram inspired Paragraph2vec (Le and Mikolov [22]) and RNN-backed methods introduced by Kiros et al. [23]. However, documents may oftentimes rather constitute a collection of words than collections of actual sentences as our use case shows. Moreover, most of these techniques require retraining when changing the word embeddings or when encountering unseen examples as De Boom et al. [21] point out.

Our method, however, can employ pre-trained word embedding models without retraining while also being able to handle documents with unknown words to a certain extent, as the following sections will show. Also note that our method does not require any labeled training pairs of documents and topics.

---

[1] Note that tf-idf-weighting topic descriptions may be useful if topic descriptions have words in common.

3 Use Case and Benchmark

The evaluation of advertisement effectiveness is a common undertaking in managing online retailers. For example, retailers may wish to analyze the content of ad serving websites that a web shop's audience is frequenting. Categorizing these websites may be helpful to oversee ad spending and to gain insights about the type of audience that the online retailer is attracting. However, sorting into arbitrary categories is usually not helpful as it hinders comparisons with established taxonomies. Hence, classifying into predefined categories is a necessity, rendering the classic LDA approach unfeasible. Labels on the other hand would be costly to acquire in such a setting since they require human input.

We evaluate the predictive performance of our topic embeddings approach against the seeded LDA approach on a dataset of URLs obtained from a German online retailer. The dataset contains approximately 380,000 records of URLs with a *.de* top-level domain, from which the vast majority are German-content websites, as a qualitative scan of the data reveals. For the word embeddings model, we follow Müller's procedure [24] to train a 300-dimensional German word2vec model. As data inputs, we use a corpora of German news articles from 2007-2013 [25] and a copy of the German Wikipedia (retrieved Dec 21, 2017).

In order to allow for a quantitative performance evaluation of both methods, we manually label a fraction of 2,488 randomly selected records, which we then randomly split into a test and holdout set in an 80:20 ratio. To get consistent labels, two authors labeled the first 500 documents together and consult each other in ambiguous cases during labelling of the remaining records.

3.1 Topic Taxonomy

We use the Google AdWords taxonomy to predefine 22 main topics [26], which are further divided into sub-, sub-sub and sub-sub-sub-categories. However, we restrict ourselves to classifying into the 22 main topics.

Table 1. Examples of the used topics and their corresponding seed words[2].

Topic	*Seed Words*
Dining & Nightlife	dining, nightlife, night club, disco, restaurant, menu, bar, drinks, DJ, party, dancing, going out
Computers & Consumer Electronics	computer, consumer electronics, pc, laptop, mouse, keyboard, game console, playstation, xbox, wii, nintendo
Food & Groceries	food, groceries, juices, soft drinks, alcohol, beverages, vegetables, fruits, pasta, rice, potatoes, supermarket, milk, cheese, eggs, butter, bread, household goods, detergents, consumer goods

[2] The originally used seed words are German and have been translated.

As the taxonomy is in English and our documents are in German, we first translate the 22 main categories and their 252 sub-categories. To create further seed words, we additionally enrich the topic descriptions with extensions as well as with some associated entities and proper names. As extensions we consider synonyms as well as words (mostly nouns) associated with a given topic, despite them not being direct translations of the original English topic. As associated entities we consider popular brands or institutions linked to certain topics. Table 1 lists seed words for three example categories. Also, we make sure to use each word in the description of only one topic and avoid words which could be associated with more than one topic altogether. In total, the final topic taxonomy consists of 421 words in the translation part, 656 words as extensions and 49 words in the associated entities section.[3]

3.2 Preprocessing: Constructing Documents from URLs

We build our documents from the word tokens inside each URL's host and path. However, in some cases, the URL is extremely short and does not provide much useful information. For example, the host may be a brand name or another non-word, combined with a very short or non-word only path. Thus, we use a web crawler to extend our URL documents by the corresponding website titles as well as description and keyword tags. We thus end up with a corpus of 384,285 documents with an average (and median) length of 24 words when building the documents from URL tokens as well as all three crawled information sources.

3.3 Grid Search: Optimization of the Hyperparameter Configuration

Our approach to classifying the URL documents is twofold and comprises of two stages: After the pre-processing stage in which the documents are build, we classify them with either one of the two methods described in section 2. Each stage includes several hyperparameters to be tuned, which influence the performance of the resulting models. To examine the influence of these parameters and to identify the best configurations for our two methods, we conduct a grid search. The specific configuration options are shown in Table 2.

Table 2. Hyperparameter space for grid search evaluation.

Area	*Options*	*Range of values*
Topics	Topic description sources	[Translations, Extensions, Entities]
	Include sub-category words	[True, False]
Documents	Source of document contents	[URL Tokens, Crawling Tokens]
	Crawling sources	[Title, Description, Keywords]
Seeded LDA	Seed Confidence	[0.2, 0.8]
	Replace non-vocabulary seeds	[True, False]
Topic Embeddings	Weight words with tf-idf	[True, False]

[3] Upon request we are happy to share the list with interested readers.

During pre-processing, the documents can be constructed from either only the words in the URL, the words obtained by the web crawler or combinations thereof. Specifically, we distinguish one URL word source, three crawling word sources (website title, description and keywords) and combinations thereof. This procedure also allows us to more closely examine the effects of including the additional information retrieved from the web crawler on our two models' predictive performances. The second part of the pre-processing stage is the aggregation of information about the topics from our taxonomy. As described in section 3.1, the taxonomy was translated and then extended by characteristic words and associated entities. For both models we test the effect of including each of these sources as well as their combinations to guide the learning process. Further, we vary the amount of information included in the topic descriptions by either combining all information from sub-categories into their respective main categories, which we want to predict on, or rather leaving out the information related to the sub-categories altogether. Also, both our classification techniques require a number of hyperparameters to be set. The seeded LDA algorithm provides a variety of choices for the seed sets: We test two options for the degree to which the initial distributions are skewed towards the seed words, comparing a very high confidence of 0.8 against a low one of 0.2. Additionally, we test for the effects of replacing seed words that are not contained in the document corpus' vocabulary with the help of the word embeddings model, by means of replacing the missing words with the most semantically similar in the corpus vocabulary. For the implementation of the topic embedding approach, we include the option of weighting the influence of each word in the document by the tf-idf score. In total, the grid search includes 384 hyperparameter combinations[4] that we evaluate on the labeled test set.

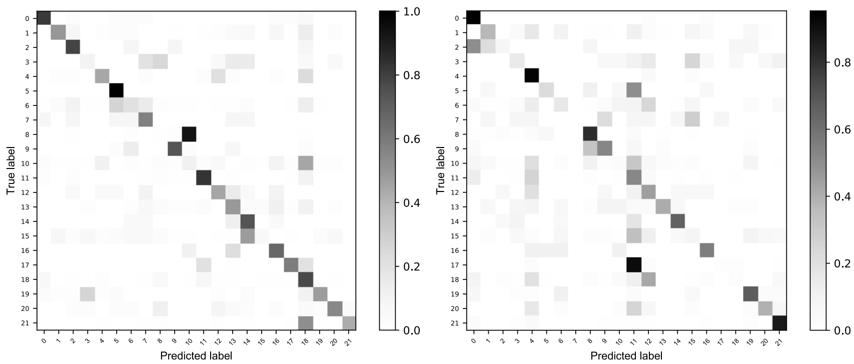


Figure 1. Confusion matrices of sensitivity-best models (left: Seeded LDA, right: Topic Embeddings). See Table 3 for topic names.

[4] We ignore combinations of topic description sources without translations since extensions and associated entities were less densely available and empty for some categories. Early trials indicated that using only extensions and/or associated entities leads to expectably poor results. Similarly, we exclude combinations of crawling sources that lack page titles because many crawled pages do not feature descriptions and/or keywords.

4 Results

The baseline for evaluating the results of the model is the effectiveness of simply predicting the most common category. To estimate this baseline score, we calculate the distribution of topics in the section of the dataset that we manually labeled. The most frequent label is "Vehicles" with a share of 18.8%. For the top-2 and top-3 scores the baseline is 30.9% and 42.3% respectively.

4.1 Seeded LDA Method

The seeded LDA approach correctly classifies a maximum of 45% of all instances in the test set in terms of sensitivity, which measures the share of correctly discovered topics, and reaches 55% in terms of precision, which indicates how useful the predicted labels are, i.e. how large the share of relevant labels is among the retrieved labels[5]. However, it shows poor performance on certain topics like “Food & Groceries”. Not using the web-crawled data decreases the performance by 13 percentage points, meaning that the sensitivity climaxes at 36% when the method only works on the tokens from the URL itself.[6] With regards to the possible combination of sources that describe the topics (translations, extensions, associated entities), it becomes clear that the extensions make a significant difference as they add 10 percentage points to the performance, while the inclusion of associated entities has no impact. Replacing seed words not found in the document corpus’ vocabulary makes no difference in our case, which is likely due to the fact that already a very large portion of the topic descriptions vocabulary is shared by the document corpus vocabulary. Further, the seeded LDA model’s top-2 and top-3 predictions cover the correct topic in 62% and 69% of all cases respectively.

4.2 Topic Embeddings Method

The topic embeddings approach reaches scores of 41% in sensitivity and 49% in precision. While it performs very well on certain topics like “Vehicles” (21), it does not do well on the “Hobbies & Leisure” (10) category and, as obvious from the confusion matrix in Figure 1b, frequently confuses "Home & Garden" (11) with "Real Estate" (17). Interestingly, the best topic embedding model relies only on data from the main categories as topic descriptions and excludes extensions and associated entities.[7] Adding the latter two or including the descriptions of sub-categories makes no difference aside from a stark increase in computation time as the number of required word to word distance calculations scales linearly with the length of the topic descriptions. The best

[5] The optimal configuration includes most document word sources (URL tokens as well as title and keywords from crawling results) and requires all seed word sources from main and sub-categories as topic descriptions.

[6] Topic descriptions again require all seed word sources and all words from main and sub-categories.

[7] The optimal model configuration further includes building documents from URL tokens as well as website keyword and title tags.

model uses tf-idf to weigh words but the influence of using tf-idf is, perhaps due to the shortness of the documents, very low. Restricting the compilation of documents from using the data obtained from the web crawler results in only a small performance drop of 3 percentage points in sensitivity and 1 percentage point in terms of precision. The topic embeddings' top-2 and top-3 scores cover the correct topic in 50% and 56% of all cases respectively, both including crawled data for document generation.

Table 3. Performance summary of the topic embedding and seeded LDA models evaluated on the test set.

			Topic Embeddings		*Seeded LDA*		
		Support	Sensitivity	Precision	Sensitivity	Precision	$\Delta_{Sensitivity}$
0	Apparel	43	0.95	0.31	0.84	0.55	0.11
1	Arts & Entertainment	119	0.37	0.69	0.50	0.79	-0.13
2	Beauty & Personal Care	14	0.07	0.09	0.79	0.29	-0.72
3	Business & Industrial	20	0.15	0.09	0.10	0.06	0.05
4	Computers & Consumer Electronics	77	0.94	0.25	0.38	0.39	0.56
5	Dining & Nightlife	18	0.22	0.36	0.94	0.25	-0.72
6	Family & Community	126	0.16	0.36	0.21	0.41	-0.05
7	Finance	14	0.00	0.00	0.64	0.30	-0.64
8	Food & Groceries	103	0.82	0.60	0.01	0.03	0.81
9	Health	49	0.53	0.52	0.73	0.77	-0.20
10	Hobbies & Leisure	237	0.03	0.17	0.12	0.21	-0.09
11	Home & Garden	159	0.53	0.21	0.80	0.71	-0.27
12	Internet & Telecom	20	0.45	0.05	0.50	0.16	-0.05
13	Jobs & Education	59	0.41	0.75	0.46	0.50	-0.05
14	Law & Government	23	0.65	0.52	0.74	0.20	-0.09
15	News, Media & Publications	90	0.08	0.15	0.06	0.23	0.02
16	Occasions & Gifts	9	0.56	0.23	0.67	0.13	-0.11
17	Real Estate	121	0.02	0.67	0.58	0.83	-0.56
18	Retailers & General Merchandise	250	0.02	0.42	0.77	0.32	-0.75
19	Sports & Fitness	67	0.67	0.48	0.48	0.82	0.19
20	Travel & Tourism	51	0.39	0.67	0.55	0.72	-0.16
21	Vehicles	367	0.86	0.91	0.41	0.99	0.45
total / average		2036	0.41	0.49	0.45	0.55	-0.04

4.3 Ensemble Learning: Combining Topic Embedding and seeded LDA

By comparing the two confusion matrices in Figure 1, one can easily discover that the two methods complement each other in parts. Thus, we train an ensemble learner whose results are depicted in Figure 2.

We combine the predictions of our two approaches in a decision tree, where the labels obtained during the grid search outlined in section 3.3 serve as features. We limit the

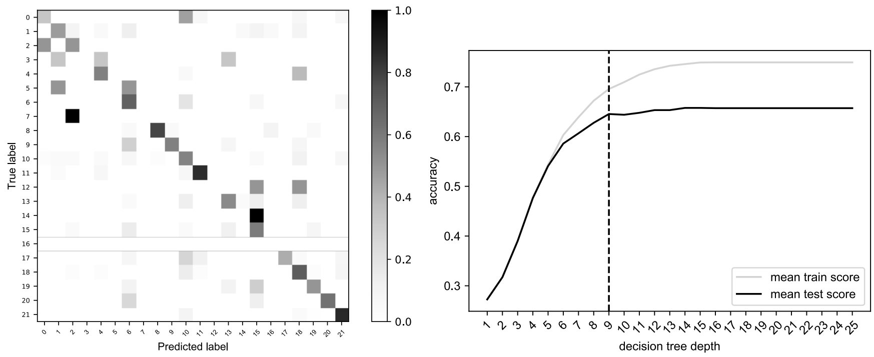


Figure 2. Left: Confusion matrix of ensemble learner evaluated on holdout-set (452 documents, none in topic 16). Right: Results of the grid search for the final ensemble learner.

resulting tree to take only two features as input due to the high computational cost of each feature, which we assume could otherwise be prohibitive in a realistic application scenario. We train a tree with all features and select the two most important ones. Not surprisingly, each of our two models contributes one feature. While the selected topic embeddings model is the best performing one described in section 4.2, the selected seeded LDA model is very similar to the best-performing one described in section 4.1 and only differs with respect to including the crawled website descriptions into the documents. Subsequently, we undertake an additional optimization endeavor with three-fold cross validation to identify the optimal tree in terms of test score, which is shaped by using entropy as a split criterion, an impurity decrease of 0 and an optimal depth of 13. We decide to further decrease the maximal depth to 9 to prevent overfitting and reduce complexity. We evaluate this final ensemble model on the holdout-set and reach scores of 65% sensitivity and 70% precision.[8]

Table 4. Performance summary for seeded LDA and topic embeddings evaluated on test-set and final ensemble learners evaluated on holdout-set.

	Options			
	Crawling		*No Crawling*	
	Sensitivity	*Precision*	*Sensitivity*	*Precision*
Seeded LDA	45%	55%	36%	49%
Topic Embeddings	41%	49%	38%	48%
Ensemble	65%	70%	50%	56%

Note: Topic word sources vary by method.

[8] See Table 4 for results of a crawl-data free ensemble learner fitted by the same procedure.

5 Discussion

We introduced a novel method for unsupervised topic modeling with predefined topics and evaluated its performance against the established seeded LDA model in a real-world application setting characterized by very short documents.

Our results emphasize that the seeded LDA method is very dependent on the length of its inputs and greatly benefits from additional data, both in terms of document length and in terms of topic description length. Since sufficiently long documents may not always be readily available as the case study demonstrates, this is one of the main shortcomings of the seeded LDA model. Nonetheless, it does perform very well when supplied with the required text volume. The topic embeddings model on the other hand does not profit from increased document length or topic descriptions. Its greatest strength is its robust performance that is almost on par with the seeded LDA model in a data-sparse environment of very short documents. The topic embeddings approach could easily be extended and possibly be refined in further research, for example by accounting for the varying performance of word embedding models depending on the word types. Rubinstein, Levi, Schwartz and Rappoport [27] show that on the one hand, word embedding models are limited with regards to learning attributive relationships like that between the words “fast” and “jet” while on the other hand they emphasize the strengths of such models regarding taxonomic relationships like that between “car” and “vehicle”. This hints at potential for performance improvements of the topic embeddings approach by only considering nouns while dropping everything else within a document. Another possible field of study lies in evaluating the effects of including bigram tokens in the model and, more far reaching, evaluating the robustness of the performance with regards to the chosen seed words. In addition, it would be interesting to compare the topic embeddings model performance against traditional classification methods in supervised settings. We suspect our method would naturally compete well against established models when training data is sparse since it has shown strong performance without learning from labeled examples, as outlined in section 4.2. However, with increasing amounts of training data, we believe it would be outperformed by traditional bag of words classifiers or other word embedding approaches like Facebook research’s fastText [28], which averages word embeddings of a document and passes the result through a softmax layer. While supervised classifiers obviously benefit from additional training data, our approach is independent of the amount of training data. It would be helpful to further investigate at which amount of data supervised learning algorithms take over.

However, one must keep in mind that both seeded LDA and the topic embeddings method are only unsupervised with regards to not requiring labeled document data. They obviously require input on the topics which one would like to classify into. Further research could possibly refine this area by investigating optimal discovery and selection strategies for seed words as well as determining which amount of information is optimal for maximum performance.

By combining both methods into an ensemble model, we showed how their individual strengths can complement each other. However, this requires one to pay the high price of labeling some data that essentially renders the whole undertaking a partly supervised method. Moreover, even the ensemble model struggles with very broad categories like "Hobbies & Leisure". Such diverse topics are problematic and further research is necessary to evaluate their handling, possibly by breaking them down into subcategories for classification and re-aggregating them afterwards.

Notable work on the combination of LDA and word vectors has been published in the past, e.g. by Nguyen et al. [29] and Moody [30]. While these combinations show strong performance, they lack our approach's ability to be steered towards a given taxonomy. Nonetheless, it would be valuable to investigate whether these existing models can be amended to allow for guiding towards given sets of categories.

To conclude, the topic embeddings model that we developed proves to be a valuable addition to current topic modeling research. Its relative independence of document length distinguishes it markedly from LDA approaches. Together with the ability to classify into given categories and its unsupervised nature, it marks one such specialized topic modeling technique that Eickhoff and Neuss [1] called for. Its strong performance on very short URL documents hints at further use cases in other areas where short documents are present, such as social media data or item descriptions in online shops. However, as language differs by context and evolves over time, one should keep in mind that the backbone of the approach, the word embeddings model, may need to be retrained on corpora of current, application-relevant text.

References

1. Eickhoff, M., Neuss, N.: Topic Modelling Methodology: Its Use in Information Systems and Other Managerial Disciplines. In: Proceedings of the 25th European Conference on Information Systems (ECIS), pp. 1327–1347. (2017)
2. Müller, O., Junglas, I., vom Brocke, J., Debortoli, S.: Utilizing Big Data Analytics for Information Systems Research: Challenges, Promises and Guidelines. Eur. J. Inf. Syst. 25, 289–302 (2016)
3. Chen, H., Zhao, J.L.: ISTopic: Understanding Information Systems Research through Topic Models. In: Thirty Sixth International Conference on Information Systems (2015).
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. J. Mach. Learn. Res. 3, 993–1022 (2003)
5. Dumais, S., Platt, J., Heckerman, D., Sahami, M.: Inductive Learning Algorithms and Representations for Text Categorization. In: Proceedings of the Seventh International Conference on Information and Knowledge Management. pp. 148–155. ACM (1998).
6. Blei, D.M., Mcauliffe, J.D.: Supervised Topic Models. Adv. Neural Inf. Process. Syst. 121–128 (2008)
7. Lacoste-Julien, S., Sha, F., Jordan, M.I.: DiscLDA: Discriminative Learning for Dimensionality Reduction and Classification. Adv. Neural Inf. Process. Syst. 897–904 (2009)
8. Jagarlamudi, J., Daumé III, H., Udupa, R.: Incorporating Lexical Priors Into Topic Models. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. pp. 204–213. Association for Computational Linguistics (2012).
9. Schwaiger, J.M., Lang, M., Johannsen, F., Leist, S.: "What Does the Customer Want to Tell Us?"-An Automated Classification Approach for Social Media Posts at SMEs. In:

Proceedings of the 25th European Conference on Information Systems (ECIS), pp. 20134–2050. (2017)
10. Murawski, M., Bick, M.: Demanded and Imparted Big Data Competences: Towards an Integrative Analysis. In: Proceedings of the 25th European Conference on Information Systems (ECIS). pp. 1375–1390. (2017)
11. Müller, O., Brocke, J.: Enriching iTunes App Store Categories via Topic Modeling. In: Thirty Fifth International Conference on Information Systems (ICIS). (2014)
12. Geva, H., Oestreicher-Singer, G., Saar-Tsechansky, M.: Using Retweets to Shape our Online Persona: A Topic Modeling Approach. In: Thirty Seventh International Conference on Information Systems (ICIS). (2016)
13. Xu, R., Chen, H., Zaho, L.J.: Topic Associations in Document Models for Ad-Hoc Information Retrieval. In: Pacific Asia Conference on Information Systems (PACIS). (2016)
14. Eickhoff, M., Wieneke, R.: Understanding Topic Models in Context : A Mixed - Methods Approach to the Meaningful Analysis of Large Document Collections. In: Proceedings of the 51 st Hawaii International Conference on System Sciences, pp. 903–912. (2018)
15. Newman, D., Chemudugunta, C., Smyth, P., Steyvers, M.: Analyzing Entities and Topics in News Articles Using Statistical Topic Models. In: International Conference on Intelligence and Security Informatics, pp. 93–104. (2006)
16. Jelodar, H.; Wang, Y.; Yuan, C.; Feng, X.: Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. Multimedia Tools and Applications, 1–43 (2017)
17. Wallach, H.M., Mimno, D., Mccallum, A.: Rethinking LDA : Why Priors Matter. In: NIPS'09 Proceedings of the 22nd International Conference on Neural Information Processing Systems, pp. 1973–1981. (2009)
18. Andrzejewski, D., Zhu, X., Craven, M.: Incorporating Domain Knowledge into Topic Modeling via Dirichlet Forest Priors. In: Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09, pp. 25–32. ACM (2009)
19. Osgood, C.E.: Semantic Differential Technique in the Comparative Study of Cultures. Am. Anthropol. 66, 171–200 (1964)
20. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. arXiv Prepr. arXiv1301.3781. (2013)
21. De Boom, C., Van Canneyt, S., Demeester, T., Dhoedt, B.: Representation Learning for Very Short Texts Using Weighted Word Embedding Aggregation. Pattern Recognit. Lett. 80, 150–156 (2016)
22. Le, Q., Mikolov, T.: Distributed Representations of Sentences and Documents. In: Proceedings of the 31st International Conference on Machine Learning. (2014)
23. Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R.S., Torralba, A., Urtasan, R., Fidler, S.: Skip-Thought Vectors. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, pp. 3294–3302. (2015)
24. Müller, A.: GermanWordEmbeddings, http://devmount.github.io/GermanWordEmbeddings/ (Accessed: 21.12.2017)
25. Koehn, Philipp: Statistical Machine Translation. (2010)
26. Google LLC. Verticals, https://developers.google.com/adwords/api/docs/appendix/verticals (Accessed 02.02.2018)
27. Rubinstein, D., Levi, E., Schwartz, R., Rappoport, A.: How Well Do Distributional Models Capture Different Types of Semantic Knowledge ? In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pp. 726–730. (2015)

28. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of Tricks for Efficient Text Classification. arXiv preprint arXiv:1607.01759. (2016)
29. Nguyen, D.Q., Billingsley, R., Du, L., Johnson, M.: Improving Topic Models with Latent Feature Word Representations. Trans. Assoc. Comput. Linguist. 3, 299–313 (2015)
30. Moody, C.E.: Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec. arXiv preprint arXiv:1605.02019. (2016)

Leveraging Unstructured Image Data for Product Quality Improvement

Oliver Nalbach[1], Maximilian Derouet[1], Anil Kumar Erappanakoppal Swamy[1], and Dirk Werth[1]

[1] AWS-Institut für digitale Produkte und Prozesse gGmbH, Saarbrücken, Deutschland
{<first_name>.<last_name>@aws-institut.de}

Abstract. Recently, traditional quality assurance methods, which often require human expertise, have been accompanied by more automated methods that use machine learning technology. These methods offer manufacturers to reduce error rates and, consequently, to increase margins as well. In particular, predictive quality assurance (PreQA) allows to minimize expenses by feeding back information from product returns and quality checks into the early product development. However, PreQA requires detailed information about previous quality problems which is not always readily available in a structured form. In this paper, we therefore discuss the potential of leveraging initially unstructured information in the form of images, taken either during quality checks or by customers when returning a product, to the end of product quality improvement. We furthermore show how this might be realized in practice using the case of fashion manufacturing as an example.

Keywords: quality assurance, image analysis, data science

1 Introduction

Manufacturers strive to avoid product defects as they have the potential to diminish margins and are detrimental to their public image. Traditionally, quality problems have been combatted by quality checks during production to avoid defective products slipping through to the customer and manual analyses of problem causes to avoid the initial bad decisions that might lead to quality problems further down the road [1]. Recently, these prevalent methods, which often require some form of human expertise, have been accompanied by more automated methods that use machine learning technology. In particular, the idea of predictive quality assurance (PreQA) [2] allows to minimize expenses due to product failures by feeding back information from product returns and quality checks into the early product development phase where then data-based predictions are made.

Predictive quality assurance requires as detailed information as possible about previous quality problems. However, quality checks are laborious and, for the case of product returns, while information is provided "for free" by the customer, it is often also very imprecise and coarse. A common situation is that of a customer taking

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

photographs of her defective product to point out the respective problems, but not using any kind of systematic annotation. This is the case for online purchases in particular as there is no sales staff to discuss issues with.

In such cases, information is available but not in a structured form so that it cannot be immediately used for machine learning purposes. To retrieve the knowledge hidden in photographs, structured information has to be mined from them. In this paper, we will discuss the potential of leveraging the initially unstructured information in the form of images, taken either during quality checks or by customers, to the end of product quality improvement, and how this might be realized in practice. The latter is demonstrated using the case of fashion manufacturing as an example.

After reviewing previous work in the next section, we summarize the predictive quality assurance approach [2] and discuss the role of unstructured data in that context (Sec. 3). In Sec. 4, we then outline an application scenario in fashion industry and show how a processing pipeline can use state-of-the-art image processing and vision methods to mine detailed information about product defects from photographs of different garments. First results presented in Sec. 5 underline the feasibility of the approach in practice. We conclude with an outlook towards future work on the topic (Sec. 6).

2 Previous Work

Existing literature overlapping with the topics of this paper comes from the domains of quality assurance, machine learning and computer vision.

2.1 Machine Learning for Quality Improvement

Applying machine learning to quality assurance is not a recent idea but has been around for a few decades. Early approaches however used to be rather passive in nature, focusing on the mere prediction of quality-related problems based on features and properties of a product and its production process. Examples include predicting whether a particular software component is error prone based on source code properties [4], or forecasting product quality in injection molding processes [5] and semiconductor manufacturing [6] given the manufacturing parameters. While such predictions can help to avoid producing low quality products they do not automatically lead to the production of higher quality products – for that, optimization approaches are necessary. E.g., given a model predicting the quality of an injection molding process in terms of properties of the resulting product, the process parameters may be tweaked automatically using a genetic algorithm [7]. Such pure optimization approaches have recently been accompanied by assistance systems which do not only allow to tweak existing products but which can already assist their users during product design and which allow the transfer of knowledge to completely new products instead of being tied to a specific process for a specific product. In particular, the predictive quality assurance architecture [2] provides a framework to implement such an assistance system. None of these existing approaches have however demonstrated how unstructured information about product quality problems can be mined and then used for an eventual quality improvement, so far.

2.2 Image Processing and Computer Vision

Research in image processing and computer vision has produced methods which are relevant to this work in two ways. First, there is a history of methods with a similar objective that use image processing to detect defects on objects either during maintenance or quality checks in production [8]. Second, for our proposed processing pipeline (Sec. 4), we transfer algorithms originally designed for different tasks such as people detection and pose estimation to serve a new, additional purpose as sub-steps in the mining of defect information from images.

For identifying defects and failures from images, there are both supervised methods, i.e. methods requiring a labeled data set to be trained on [9-11], and unsupervised methods [12-14], purely based on detecting statistical irregularities. Some aim at generic anomaly detection [14] while others have fixed application domains ranging from wood boards [13] to steel rails [11] to textiles [10]. The vast majority of those methods assumes their input images to be of a very regular nature, only showing a relevant patch of an object's surface texture, not the whole corresponding objects. This requires images to be taken under pre-defined conditions either automatically using calibrated equipment at a production site, e.g., at a fixed distance, angle and under pre-defined lighting conditions, or manually by an instructed expert. Consequently those approaches cannot easily be transferred to the much more irregular photographs taken by customers using varying hardware and showing different views of a full object or even just showing drastically varying complex objects such as different garments.

We post-pone the discussion of related work originally developed in other contexts, which we propose for different processing steps, to Sec. 4 where we will introduce our defect mining methodology.

3 Background

In this section we first summarize the main ideas behind the predictive quality assurance architecture [2] before going on to discuss the role and benefits that image data may provide in this context.

3.1 Predictive Quality Assurance

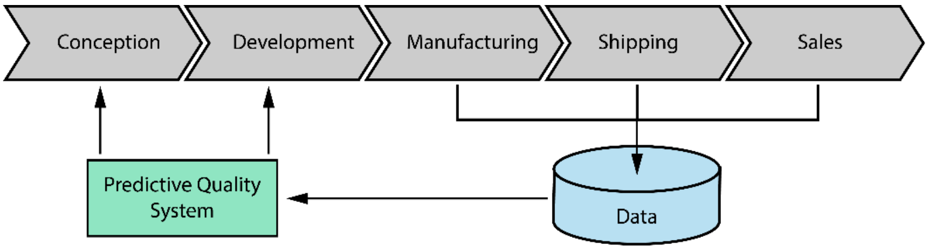


Figure 1. Predictive quality assurance: Data about products and the problems they exhibited previously is collected throughout the depicted process from conception to sales. The predictive quality system analyzes this data to gain insights which are fed back into the early process steps.

Predictive quality assurance (PreQA) uses historical data about manufactured goods and the defects they exhibited to eventually increase the quality of future products. At the core of the method are classification algorithms which are trained to predict the expected occurrence rates for different types of product defects, given a structured description of product features as input. In particular, predictions can also be made given only a subset of all possible features so that the method can already be applied in early stages of product development where not all existing features have been specified yet (Fig. 1). Using these predictions, issues such as bad product designs can be identified early-on. An integrated assistance component then allows to correct adverse decisions made about the product by suggesting alternative options with significantly lower expected probabilities of failure.

3.2 The Value of Image Data for Quality Assurance and Improvement

As PreQA is based on classification algorithms, the availability of a high quality dataset to train these algorithms on is crucial to the method's success. This not only concerns the level of detail found in the specifications of existing products, which should comprise a large set of expressive features, but furthermore especially the descriptions of previous cases of defect.

Using the case of fashion industry as an example, product defects can be of drastically varying nature: holes, stains, fragile seams, broken zippers, lost buttons, bleached colors or strong pilling all are reasons for a customer to return a piece of clothing. But they have very different causes. A bad material mix may affect pilling behavior but arguably has little influence on oil stains caused by a production machine. A weak thread can lead to loose buttons but will not cause problems with a zipper. Thus, if different types of failure are clustered together into a single class of defect, valuable potential for a detailed analysis and a subsequent reliable prediction and failure avoidance is lost irretrievably.

Unfortunately, when returning an item bought from a store due to a defect, customers are usually required to specify their reasons for doing so only in a very coarse way. This is because customers, on one hand, cannot be expected to analyze the specific nature of product failures as they are no domain experts and, on the other hand, because it would take additional time and lead to increased frustration at the customer's end. While the provided coarse defect information can give first insights into problems connected to specific products the imprecision makes it difficult to connect particular defects to individual product features using machine learning. This is reflected in the fact that the prediction confidence of PreQA is usually lower for unspecific defect labels while being higher for more specific ones [2].

Attaching unstructured data in the form of images to product return cases offers a way to mitigate this problem: A picture is worth a thousand words. Using photographs, we can generate detailed descriptions of situations in the fraction of a second. Taking a picture of her defective item takes the customer less effort than filling out a detailed return form by selecting appropriate defect descriptions from dropdown lists but still encodes virtually all information about the occurred problem into a compact visual description. The same applies to quality checks of final products performed in companies. While in this case the staff is specialized and experienced in assessing defects, just taking a photograph could significantly decrease the time required for checks which would allow a denser sampling of the full production yield given the same time budget and increase the quality of the items sold by itself.

3.3 Manual Visual Inspections in Fashion Companies

An automated analysis of images showing defective products should ideally produce output which is identical to the one expected from a human expert assessing the case manually. We will thus briefly describe the way visual inspections proceed for the exemplary case we are dealing with in this paper, the fashion industry.

While there are some industrial norms for quality checks in the form of ISO or DIN standards in clothing industry they mainly concern apparel which is used in safety-critical contexts. When it comes to quality checks of casual clothing, each company defines its own inspection protocols to follow. Nevertheless, there are certain best practices performed in most companies.

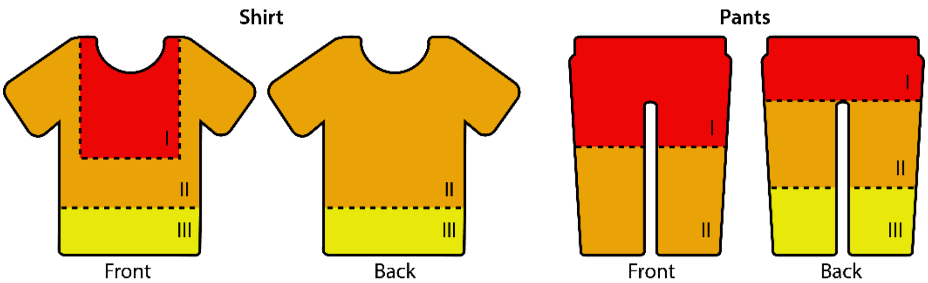


Figure 2. Typical defect zones used in visual garment inspections, according to [19]

First, a list of possible defects is defined. These potential quality flaws are then considered one by one during visual inspections. Second, garments are commonly subdivided into different zones called e.g., zone A to C or I to III, as depicted in Fig. 2.

These zones often correspond to different levels of visibility, and consequently also severity, of defects. Usually, the highest severity is assigned to parts visible when presenting the product in a store (Fig. 2, zones I), the second highest to areas more visible when wearing it (Fig. 2, zones II) and so on. Third, each company defines combinations of defect types and defect locations, in terms of zones, which lead to the rejection of a piece of clothing or its re-use in outlet stores, respectively.

4 A Processing Pipeline for Mining Textile Defects from Images

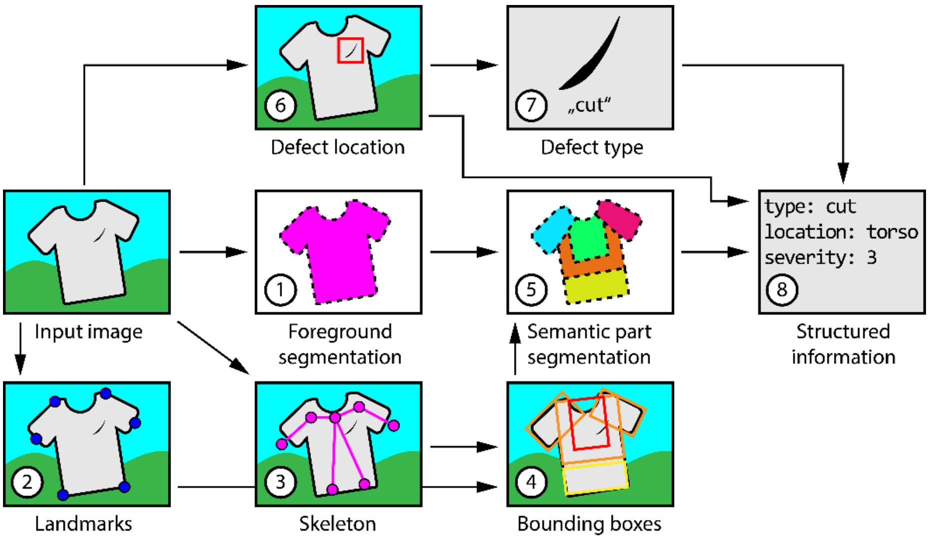


Figure 3. Image processing pipeline to mine details about product defects from images for the case of fashion industry. Steps are labelled by the order in which they are described.

We will now outline how computer vision techniques can be used to compose a processing pipeline to mine textile defects from photographs of pieces of clothing. An overview is shown in Fig. 3. It is inspired by the best practice approaches used in companies (Sec. 3.3) on one hand and the requirements set by machine learning algorithms on the other hand. The input to the pipeline is a photograph of a defective garment. No additional annotation is required. The output of the pipeline consists of a defect description regarding three properties: the type of the defect, its location with respect to the garment and a severity rating depending on the type and location which can be used to decide how to proceed further with the specific item. The pipeline consists of seven processing steps. For each of these steps, we will describe which output they compute, why they are necessary and which existing methods can be used to implement them.

4.1 Foreground Segmentation

Customers and customer care staff may take photographs of defective products in front of varying backgrounds. The first processing step therefore consists of segmenting the actual object of interest from that background (Fig. 3, step 1). In technical terms, each pixel of the original image is labeled as either belonging to the piece of clothing or to the background. This step is necessary for computing a more detailed segmentation of the object into different zones later on (Fig. 3, step 5) and also provides the precise areas of the image where defects can be expected to be localized later on (Fig. 3, step 6). There are several approaches which deal with this problem in different contexts, for example from the domains of image retrieval [15] or semantic scene description [16]. Some of them work best on photographs in which the clothing is worn by a person as they are based on detecting human features in a pre-processing step [15] while others are solely based on segmentation, such as the method by Borras et al. [16], and could therefore be applied in a more general setting.

4.2 Landmark Detection

In the second step, different semantic landmarks are detected (Fig. 3, step 2). Those are important image locations corresponding, e.g., to the end points of sleeves or of the collar. The output of this computation step is a list of detected landmarks – which may also correspond to only a subset of all known types of landmarks, depending on the type of garment and the specific view – and of their corresponding positions with respect to the image. The landmarks allow for a subsequent computation of 2D bounding geometry that then serves as guidance for the full semantic segmentation of the object (Fig. 3, step 5). Fashion landmark detection constitutes a very new area of research and has been introduced only recently by Liu et al. [17]. Their method employs deep neural networks and is robust regarding whether an image contains a person or not and whether the full item is depicted or only part of it (cf. Fig 4, second column).

4.3 Skeleton Fitting

In addition to the landmarks computed in step 2 which correspond to different extremal points of the pieces of clothing, further important key points can be derived using so-called pose estimation algorithms which try to fit an approximate skeleton to person depicted in an image (Fig. 3, step 3). These skeletons consist of a number of detected joints, connected by lines corresponding to different limbs (Fig. 4, third column). While those methods are originally used to determine the poses of humans in photographs or videos they often work surprisingly well for images without people in them (cf. Fig. 4, row 1) as the underlying neural networks have learned to use clothing as an important feature. Specifically, we employ OpenPose [20] to gather the additional pose information when possible. This framework outputs a number of human joint locations with respect to the input image. When particular joints cannot be detected reliably, only the subset which is detected confidently is output.

4.4 Fitting of Bounding Boxes

In the fourth processing step, the shape of the garment is approximated coarsely by a collection of two-dimensional boxes (Fig. 3, step 4). Fitting such so-called geometric proxies greatly simplifies the definition of rules to use by the semantic segmentation step following next. For example, in the case of an upper-body garment such as a shirt or jacket we first compute a box covering the torso region by fitting it to contain the four landmarks corresponding to the collar and the bottom of the torso. Boxes corresponding to the different defect zones within the torso region (Fig. 2) can then be derived from it: The zone III region typically corresponds to the lower 20% of the torso region while zone I covers 50% of the torso height and 70% of its width as it corresponds to the portion of the garment visible when folded and presented in a shop. To determine the height of the sleeves' bounding boxes we use the sleeve-ends landmarks in combination with the detected shoulder joints from step 3, if applicable, and fall back to using the collar landmarks from step 2 otherwise. The width of the sleeve boxes is derived from the horizontal distance between the collar landmarks. Results for this step are shown in the fourth column of Fig. 4.

4.5 Semantic Segmentation

Given the coarse shape approximation from step 4, as well as a mask segmenting the item of clothing from the background (step 1), the derivation of a complete segmentation (Fig. 3, step 5) can be defined using a set of processing rules: Each pixel is assigned to the semantic part corresponding to the closest box. In case a pixel is contained in multiple boxes, precedence is given to boxes corresponding to higher severity and main boxes are preferred over boxes corresponding to sleeves or legs. The resulting pixel-wise labelling is used to derive the location of the defect can then be used to rate its severity. While there are previous methods to solve this segmentation problem sometimes also referred to as clothes parsing [18] many of them strictly require the garments to be worn by a person [15, 18] and none take the typical zone assignment used in textile quality checks into account.

4.6 Anomaly Detection

Before we can classify textile defects into different categories, it is useful to first determine the image region where the defect is located (Fig. 3, step 6). The classification following in step 7 can then be applied to a cropped image showing only that region which significantly simplifies the problem as the cropped images to be classified are more regular. Assuming that most items of clothing are largely homogeneous regarding the fabric they are made of, we can employ unsupervised learning methods [12-14] which detect statistical irregularities and output a bounding box around the anomaly which can be used to determine the region to crop. By limiting the detection to the garment, using the segmentation from step 1, we can avoid irregularities being detected outside of the actual product.

4.7 Defect Classification

In the final image processing step (Fig. 3, step 7) we have to classify defects on an otherwise homogeneous image showing a textile patch. This problem has been previously tackled to automate visual quality checks of (raw) fabrics which constitute the feedstock for apparel manufacturing. There even is a standard benchmark, the TILDA dataset [21], which contains several hundred images of fabric patches with different defects and has been commonly used to assess different classification methods [22-24]. Those conventional methods are based on a two-step approach where first statistical features are computed in order to reduce the dimension of the input data (a patch of just 512 by 512 pixels with 3 color channels corresponds to a vector of 786,432 components) and then a classification only based on these features is performed.

However, recently, deep artificial neural network structures based on convolutional layers as the main computational units, called (deep) Convolutional Neural Networks (CNNs), have become very popular for various image processing and computer vision tasks [25]. In particular for classification tasks such as the annual ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [27] they have outperformed more traditional methods by a wide margin. We therefore compare their performance for the task of textile defect classification to that of existing methods [22-24] in Sec. 5.2.

4.8 Data Fusion

After we have gathered information about the different possible defect zones with respect to the piece of clothing (Step 5), the location of the defect (Step 6) and its type (Step 7), we can output corresponding structured information (Fig. 3, Step 8) which can then be used to train more precise classifiers for PreQA.

5 Results and Discussion

In this section we present initial results for several steps of our processing pipeline, specifically, the computation of bounding boxes and the final classification step.

5.1 Clothes Segmentation

Using the publicly available implementations of OpenPose [20] as well as of the fashion landmark detection of Liu et al. [17] we implemented steps 2-4 of our pipeline. Fig. 4 demonstrates the output of these steps for the case of four upper-body garments. Results are shown for both images with (Fig. 4, rows 2+3) and without humans (rows 1+4) and for the cases of both successful (rows 1+2) and unsuccessful (rows 3+4) pose detection.

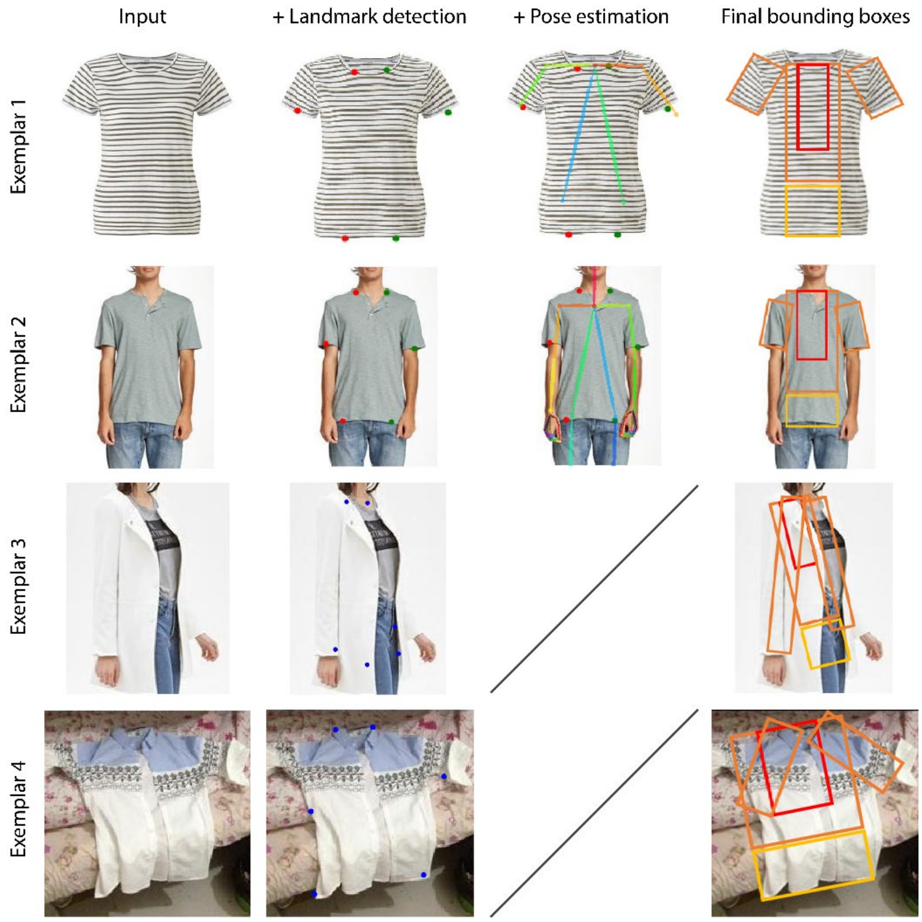


Figure 4. Pipeline input (first column) and outputs of processing steps 2 to 4 (second to fourth column) of our pipeline (Fig. 3). For rows three and four no full skeleton estimation was possible and the fallback approach (Sec 4.4) was used. The sample images have been taken from the fashion landmark detection benchmark dataset [17].

Table 1. Metrics achieved on the test set. As we deal with a multiclass classification problem the values have been averaged across the (evenly balanced) classes.

Metric	*Value*
Accuracy	90.1 %
Precision	87.8 %
Recall	87.6 %

5.2 Textile Defect Classification using Deep Learning

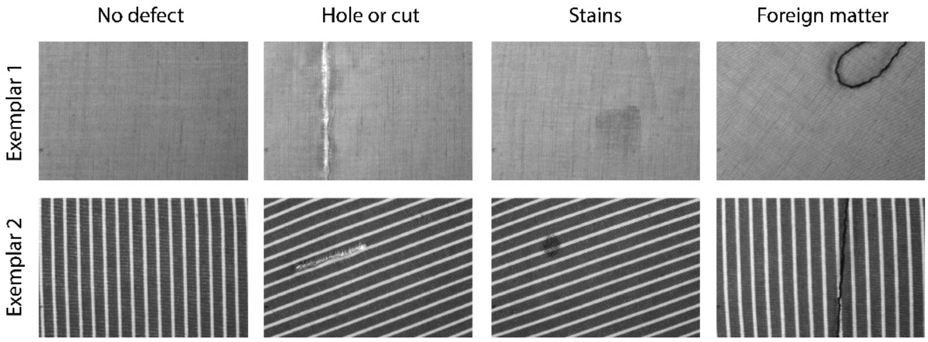


Figure 5. Sample images taken from the TILDA dataset [21] for four of the eight available classes (columns) and two of the included representatives (rows).

To evaluate the use of deep learning for textile defect classification, compared several popular network architectures from the ILSVRC contest [27], namely AlexNet [31], SqueezeNet [30] and GoogleNet [28]. We randomly split the TILDA dataset (cf. Fig. 5 for samples) into 90% of training and 10% of test images. The original balanced ratio of the classes (each class contributes 12.5%) was maintained. As TILDA only comprises about 3200 images, which is several orders of magnitude less than for most common deep learning datasets, we augmented our training data by adding mirrored and rotated copies of the original images. While TILDA contains high resolution images in a 4:3 format, all compared network architectures expect quadratic input images of lower resolution. We therefore padded images with white and then down-scaled them to the appropriate input resolution for each network. We first trained the candidate networks using the Adam solver [29] until their classification accuracy converged. As GoogleNet reached the highest initial accuracy, we selected it for further finetuning and increased the original drop-out rate to counteract potential overfitting to the small dataset, achieving an accuracy of 90% on the test set. A detailed evaluation using different common metrics is given in Table 1. This is on-par with values reported for competing methods [22] for which however no details, about whether a separate test set was used and which size it had, are given.

5.3 Discussion

The defect mining pipeline of Sec. 4 comprises eight steps, all of which constitute challenging problems by themselves. In the following, we discuss the current limitations of the most important steps and how these might be overcome in the future.

Fashion Landmark Detection While the method by Liu et al. applied by us for the problem of fashion landmark detection already achieves detection rates of about 70% when accepting only small misplacements of landmarks [17] this particular detection problem is a fairly new area of research. Improved, more robust methods can be expected to appear in the near future. This is also crucial for our defect mining approach as we use fashion landmarks as main guidance when constructing bounding geometry for different clothing zones.

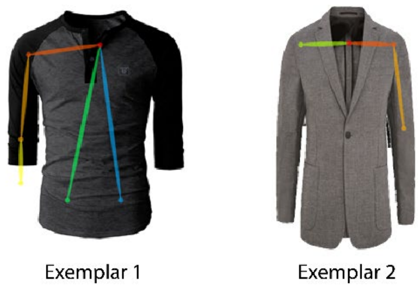


Figure 6. Two cases in which skeleton fitting failed for clothing-only images. The images shown are part of the fashion landmark detection benchmark dataset [17].

Skeleton Fitting We use skeleton fitting with OpenPose [20] as an additional input to help improve the subdivision of garments. While skeletons can sometimes also be derived for images where an item of clothing is not worn by a human (Fig. 4, first row), this cannot be expected to work in all cases. Fig. 6 shows two examples where skeleton fitting fails partially. Currently, we completely ignore the pose information in these cases. In future work, it might be useful to still extract available information for those parts of the skeleton that could be fit with significant confidence. For instance, for the left exemplar in Fig. 6, all joints but the ones corresponding to the left arm have been placed correctly and could have contributed information for fitting bounding boxes.

Bounding Boxes At the moment, we use simple boxes as the geometric proxies for different zones on pieces of clothing. This is sufficient in many cases but there are cases where a box provides a poor fit to the actual shape of the object, e.g., in cases where the sleeves of a shirt are not shown straight but bent. In such situations the fitting could be improved either by using multiple boxes, one for the upper and one for the lower part of a sleeve, or by replacing the boxes by more complex shapes such as two-dimensional generalized cylinders [26].

False Positive Anomalies We propose to use anomaly detection to spot locations of defects with respect to pieces of clothing. This assumes that their material is mostly homogeneous. While advanced detection methods also work for textured image regions, modern fashion sometimes uses defects such as holes or abrasion as design elements. Even for humans it is not always possible to decide with certainty whether a particular defect has been created voluntarily or not, unless additional exemplars of the same product are available to compare to. Consequently, it is debatable whether artificial intelligence can be expected to perform this kind of distinction more reliably.

6 Conclusion and Future Work

In this paper, we have discussed the importance of detailed training data when applying machine learning for quality improvement of manufactured goods by means of predictive quality assurance and demonstrated how image data can become a significant part in it. The latter was achieved by presenting the first pipeline to mine detailed structured defect information from images of defective products. While this pipeline is currently still of prototypical nature, preliminary results for several of the computational steps show its potential. Future work will see a full implementation of the described approach as well as a transfer to additional application scenarios in different industries, such as furniture manufacturing, where analogous methods to realize the individual processing steps, e.g., segmentation or anomaly detection, already exist within different contexts in computer vision research.

References

1. Morlock, S.: Qualitätsgesicherte Produktentwicklung in der Bekleidungsindustrie, https://www.hohenstein.de/de/inline/pressrelease_45248.xhtml?excludeId=45248 (Accessed: 30.08.2018)
2. Nalbach, O., Linn, C., Derouet, M., Werth, D.: Predictive Quality: Towards a New Understanding of Quality Assurance Using Machine Learning Tools. In: 21st International Conference on Business Information Systems. Springer, Cham. (2018)
3. Ng, H.F.: Automatic thresholding for defect detection. Pattern recognition letters, vol. 27(14) (2006)
4. Lanubile, F., Visaggio, G.: Evaluating predictive quality models derived from software measures: lessons learned. Journal of Systems and Software, vol. 38(3) (1997)
5. Ali, I.G., Chen, Y.-T.: Design quality and robustness with neural networks. IEEE Transactions on Neural Networks, vol. 10(6) (1999)
6. Chen, W.-C., Lee, A.H.I., Deng, W.-J., Liu, K.-Y.: The implementation of neural network for semiconductor PECVD process. Expert Systems with Applications, vol. 32(4) (2007)
7. Shen, C., Wang, L., Li, Q.: Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method. Journal of Materials Processing Technology, vol. 183(2-3) (2007)
8. Xie, X.: A review of recent advances in surface defect detection using texture analysis techniques. ELCVIA Electronic Letters on Computer Vision and Image Analysis, vol. 7(3) (2008)
9. Wiltschi, K., Pinz, A., Lindeberg, T.: An automatic assessment scheme for steel quality inspection. Machine Vision and Applications, vol. 12(3) (2000)
10. Kumar, A. Neural network based detection of local textile defects. Pattern Recognition, vol. 36(7) (2003)
11. Mandriota, C., Nitti, M., Ancona, N., Stella, E., Distante, A.: Filter-based feature selection for rail defect detection. Machine Vision and Applications, vol. 15(4) (2004)
12. Escofet, J., Navarro, R. F., Garcia-Verela, M. S. M., Pladellorens, J. M.: Detection of local defects in textile webs using Gabor filters. Optical Engineering, vol. 37(8) (1998)
13. Kim, C. W., Koivo, A. J.: Hierarchical classification of surface defects on dusty wood boards. In: Pattern Recognition Letters, vol. 15(7), pp. 713–721. (1994)

14. Tsa, D. M., Wu, S. K.: Automated surface inspection using Gabor filters. In: The International Journal of Advanced Manufacturing Technology, vol. 16(7) (2000)
15. Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., Berg, T. L.: Parsing clothing in fashion photographs. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3570–3577. IEEE (2012)
16. Borras, A., Tous, F., Lladós, J., Vanrell, M.: High-level clothes description based on colour-texture and structural features. In: Iberian Conference on Pattern Recognition and Image Analysis, pp. 108–116. Springer (2003)
17. Liu, Z., Yan, S., Luo, P., Wang, X., Tang, X.: Fashion landmark detection in the wild. In: European Conference on Computer Vision, pp. 229–245. Springer (2016)
18. Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., Berg, T. L.: Parsing clothing in fashion photographs. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3570–3577. IEEE (2012)
19. Fashion2Apparel, http://fashion2apparel.blogspot.com/2017/01/defects-zone-apparel-inspection.html (Accessed: 29.08.2018)
20. Cao, Z., Simon, T., Wei, S. E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. arXiv preprint (2016)
21. Technische Universität Hamburg-Harburg. Ein Referenzdatensatz zur Evaluierung von Sichtprüfungsverfahren für Textiloberflächen. Technical Report (1996)
22. Murino, V., Bicego, M., Rossi, I. A.: Statistical classification of raw textile defects. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, pp. 311–314. IEEE (2004)
23. Schael, M.: Texture defect detection using invariant textural features. In: Joint Pattern Recognition Symposium, pp. 17–24. Springer, Berlin (2001)
24. Sezer, O. G., Erçil, A., Ertuzun, A.: Using perceptual relation of regularity and anisotropy in the texture with independent component model for defect detection. In: Pattern Recognition, 40(1), pp. 121–133 (2007)
25. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012).
26. Fang, X. S.: The extruded generalized cylinder: A deformable model for object recovery. In: 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 174-181. IEEE (1994)
27. Russakovsky, O., Deng J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg A. C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. IJCV (2015)
28. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)
29. Kingma, Diederik P., and Jimmy Ba.: Adam: A method for stochastic optimization. arXiv preprint (2014)
30. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., Keutzer, K. : Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. arXiv preprint (2016)
31. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 1097-1105. (2012)

Decision Support for Real Estate Investors: Improving Real Estate Valuation with 3D City Models and Points of Interest

Markus Rosenfelder[1], Gunther Gust[1], and Dirk Neumann[1]

[1] Albert-Ludwigs-Universität Freiburg, Information Systems Research, Freiburg, Germany
{markus.rosenfelder,gunther.gust,dirk.neumann}@is.uni-freiburg.de

Abstract. Precise estimates of attainable revenue are essential for real estate investments. The large-scale availability of new data sources, combined with analytics, can help to improve price estimates in the real estate market, which is characterized by non-transparency. In this paper, we show how 3D city models and points of interest improve real estate valuations. We propose a hedonic pricing model with more than 100 features consisting of nearby points of interest as well as 3D digitized representations of buildings. We evaluate our model by means of a large showcase study in Berlin. The analysis shows that the combination of 3D data and points of interest reduces the prediction error considerably in comparison to standard approaches.

Keywords: Real estate valuation, 3D data, points of interest, decision support, data science

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Knowledge Discovery from CVs:
A Topic Modeling Procedure

Alexander Schiller

University of Regensburg, Department of Management Information Systems, Regensburg, Germany
alexander.schiller@wiwi.uni-regensburg.de

Abstract. With a huge number of CVs available online, recruiting via the web has become an integral part of human resource management for companies. Automated text mining methods can be used to analyze large databases containing CVs. We present a topic modeling procedure consisting of five steps with the aim of identifying competences in CVs in an automated manner. Both the procedure and its exemplary application to CVs from IT experts are described in detail. The specific characteristics of CVs are considered in each step for optimal results. The exemplary application suggests that clearly interpretable topics describing fine-grained competences (e.g., Java programming, web design) can be discovered. This information can be used to rapidly assess the contents of a CV, categorize CVs and identify candidates for job offers. Furthermore, a topic-based search technique is evaluated to provide helpful decision support.

Keywords: Text Mining, Topic Modeling, Latent Dirichlet Allocation, Human Resource Management

1 Introduction

Acquiring the right personnel is one of the most critical success factors for companies [1, 2]. In this area, recruiting via the web has gained significant importance over the last years [3]. It is not only of practical interest, but has also received much scientific attention (cf., e.g., [4, 5]). Opportunities are, for example, provided by well-known professional online social networks such as *LinkedIn* and *XING*, which are becoming highly popular. For instance, as of Q2 2018, *LinkedIn* has more than 562 million members [6]. Recruiting via the web is further made possible by CVs provided not only on these networks, but also on private homepages and websites specializing in making available a wide range of CVs (e.g., *Indeed*, *CareerBuilder*, *Monster*). Overall, a huge number of CVs can be acquired online. Based on these CVs, companies have the prospect to identify promising job candidates and to conduct proactive recruiting.

However, to capitalize on this potential, a very large amount of semi- and unstructured data needs to be analyzed. While approaches for a manual analysis of document collections exist [7], this task becomes too time-consuming for large collections of complex documents (such as CVs) [8]. This issue is addressed by

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

automated text mining methods, which have already been used successfully for human resource management (HRM) [9, 10]. In particular, topic modeling approaches such as Latent Dirichlet Allocation (LDA) are promising in this application context. They are able to discover the hidden thematic structure present in a document collection [11]. Thus, they should be able to extract key information from CVs. More precisely, high-quality, fine-grained topics in a specialized topic model should represent skills, abilities, knowledge and work expertise (in the following subsumed by "competences" as in [12]). This information can then be used to, for instance, rapidly assess the contents of a CV, categorize CVs and identify candidates for job offers. Topic models offer unique advantages compared to a keyword search on existing platforms. However, research has not yet discussed the application of topic modeling to CVs, leaving open crucial issues (cf. Section 2.2). This paper thus focuses on the following research question: *How can topic modeling be used to discover knowledge from CVs?*

The remainder of the paper is structured as follows. In the next section, we discuss the problem context as well as related work and the research gap. In Section 3, we propose a procedure for knowledge discovery from CVs. Section 4 contains an application of the procedure and an evaluation of the results. Finally, we provide conclusions, limitations and directions for future research.

2 Background

In this section, we first briefly introduce topic modeling, LDA and evaluation methods for topic models. Then, we give an overview of related literature and discuss the research gap.

2.1 Problem Context

Topic modeling approaches aim to discover the latent thematic structure in a document collection and to identify thematically similar documents [11]. A topic model consists of a number of topics, each represented by terms strongly associated to the topic. Recently, probabilistic approaches have been highly popular. Here, topics can be seen as probability distributions over terms and documents as probability distributions over topics. In this paper, we focus on LDA [13], a probabilistic approach not relying on any kind of training data. It is the most widely-applied topic modeling approach [14]. Distributions are calculated using sampling or optimization procedures which take into account term-document-frequencies [15]. LDA is based on the bag-of-words model (i.e., the order of words in documents is ignored). This makes it particularly suitable for CVs, which often are formulated in note form instead of continuous text.

While an evaluation of topic models by humans is considered the gold standard [16], the required time effort has sparked the need for automated evaluations, especially for testing a large number of pre-processing and parameter configurations. Many criteria and methods have been proposed, discussing, for instance, the similarity [17], stability [14] or semantic coherence [18–20] of topics. While a negative correlation to human interpretability has been reported for some methods [16], semantic coherence has been shown to provide assessments of high quality [8, 19–21]. It is often calculated based on normalized pointwise mutual information (NPMI) [19, 21]. The idea is that a topic is of higher quality when the terms strongly associated to the topic often occur in close proximity in a text corpus. Following the discussion above, measuring semantic coherence by NPMI is used in this paper for assessing various topic model configurations before the final topic model is humanly interpreted.

2.2 Related Work and Research Gap

Topic modeling and in particular LDA is widely applicable to a large range of contexts and has been successfully employed to, for instance, consumer good reviews [8], research articles [22], hotel critiques [23] and even in BPM [24]. A tutorial for applying LDA in IS in general has been proposed as well [8]. However, the high degree of abstraction and flexibility also come at a price: An adaption to the application context is necessary to provide proper results. Despite much existing work to use text mining in HRM [9, 10], a literature search revealed that little of it has addressed topic modeling.

A notable exception is a work suggesting topic modeling for job offers [4]. The objective was to identify competences of importance in the construction industry. This research is similar to ours in the sense that a topic modeling approach was used for this task. However, the aim differs, because instead of CVs, job offers have been analyzed. Furthermore, no procedure for knowledge discovery is described.

In further research, *LinkedIn* profiles of BPM professionals are examined via topic modeling to investigate the role of gender in BPM [12]. While here, a topic modeling approach is applied to documents which are similar to CVs, the aim of the research is completely different to ours. Usability for recruiting is not discussed and no procedure for knowledge discovery is presented.

To sum up, none of the existing works has proposed a procedure for knowledge discovery from CVs using topic modeling. Thus, following existing literature in this regard leads to multiple crucial issues, which constitutes the research gap addressed by this paper: First, not considering the type and characteristics of documents to be analyzed produces non-optimal results. For instance, this is due to generic text pre-processing (e.g., in the case of CVs, no removal of author's contact details). Second, existing approaches cannot even be readily applied, as essential steps are not described. For example, the acquisition of CVs and their general pre-processing is not discussed in existing topic modeling literature. Finally, the goals of applying topic modeling to CVs are not considered in existing works. This means that, critically, it remains unclear how to use the topic modeling results for actual benefit in HRM (e.g., for recruiting).

3 Knowledge Discovery from CVs

Subsequently, our procedure for knowledge discovery from CVs is presented. Figure 1 illustrates the five steps of the procedure, adapted from the process for topic modeling in general IS as proposed in [8]. After the initial acquisition of CVs (i), general pre-processing (ii) and text pre-processing (iii) are required. Only then, the preprocessed CVs can be analyzed by applying LDA (iv). Finally, the results of the application can be interpreted and used (v). The five steps are described in detail in the following.

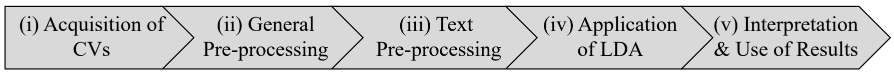


Figure 1. Steps for knowledge discovery from CVs

3.1 Acquisition of CVs

With unstructured data proliferating on the web, many options for acquiring CVs are available. The most prominent ones are utilizing (a) professional social networks such as *LinkedIn*, (b) specialized portals such as *Indeed* and (c) a web crawler.

(a): Professional social networks offer members the opportunity to present themselves to companies and recruiters via disclosing information on their profile. This includes, in particular, past work experience, education, skills, abilities, publications and interests as well as contact information. The information can be extracted from the profiles to generate CV-like documents. Internal and external tools for extraction are readily available for the most common professional social networks (e.g., *LinkedIn*, *XING*). For example, *LinkedIn* itself offers a native functionality to export member profiles as CVs in PDF format. Moreover, these portals allow members to directly upload their CVs, which can then be accessed and stored.

(b): Many job portals (e.g., *Indeed*, *CareerBuilder*, *Monster*) encourage their users to post their CV. These portals offer a (keyword) search engine which can be used to obtain CVs. For example, in Q2 2018, an exemplary search for CVs with "Data Scientist" as job title in New York City produced over 1,100 hits on *Indeed*. CVs resulting from a search can be accessed and, subsequently, stored in PDF format.

(c): Another opportunity for the acquisition of publicly available CVs is the use of a web crawler. A web crawler is an automated program able to navigate the web and store relevant information. Specifically, such a web crawler can be fed with desired search terms and programmed to find and store PDFs including these search terms. This allows the acquisition of CVs from the general searchable web. In particular, also CVs available on private homepages can be found and stored.

3.2 General Pre-processing

Once a sufficient quantity of PDFs containing CVs has been obtained, a general pre-processing of this collection of data is required to ready the collection for text pre-processing and further analyses. Depending on the way the PDFs were obtained, the common challenges (C1) or (C2) may need to be resolved:

(C1) The language of the documents does not match, causing problems for many text pre-processing routines. To address this issue, an approach for automatic language identification [25, 26] can be used. A high quality of automatic language identification can be achieved as CVs contain a substantial number of words.

(C2) The collection of data does not exclusively consist of CVs (e.g., when the PDFs were obtained using a web crawler). Obviously, this issue can lead to unsatisfactory results in the later stages of analysis, for instance, when a job offer instead of a CV is erroneously assessed to be the optimal match for a search query. A human is able to almost instantly decide whether a PDF is a CV or a non-CV with a very high degree of confidence. However, a manual distinction of CVs and non-CVs may still not be promising due to the substantial time effort required for assessing a large collection of data by hand. Thus, automated methods such as classification algorithms can be used.

In any case, PDFs should be converted to a more manageable format (such as TXT) for further analyses, and be fitted in a database for storage. As suggested in [8], an exploratory data analysis may be performed to detect possible data quality issues and to obtain a general understanding of the data to be analyzed.

3.3 Text Pre-processing

Meaningful text pre-processing before the application of a topic modeling approach is of high importance [8]. This is particularly the case for CVs, which contain many terms or even whole components irrelevant to the envisioned goal of knowledge discovery. There are simple and well-known pre-processing routines which are accepted to be valuable for (almost) all kinds of texts. Common examples are the removal of formatting tags and special characters, tokenization (i.e., splitting up documents into words), lowercasing and the removal of words occurring only in few documents. These routines can be looked up in renowned sources [27, 28]; in the following, we discuss routines which possess special characteristics with regard to CVs more explicitly.

N-gram-Creation. N-grams (expressions consisting of two or more words) instead of single words can be considered for further text analysis. For instance, many competence descriptions (e.g., ability in software such as Visual Studio) in CVs contain multiple words. Thus, a thorough creation of n-grams may be of importance. However, care needs to be taken because many skill descriptions in CVs are often used together but are not a real expression. For example, disclosing language skills in English and Spanish is common, which might lead to an incorrect 2-gram "English Spanish".

Stop Word Removal. Words that occur frequently, but are uninformative and decrease the quality and interpretability of topics need to be removed. To achieve this goal in the context of CVs, multiple types of words have to be eliminated. The first type includes general language-specific stop words, which usually are words that have only a grammatical or syntactical function such as propositions. The second type are CV-specific stop words, which are words commonly occurring in all kinds of CVs (e.g., "resume", "name"). The third type are stop words specific for the CV database at hand. To identify these stop words, word frequency lists can be used [27]. Finally, numbers may or may not also be seen as "stop words". When competences are to be modeled as topics, numbers tend to obscure the results; thus, they should also be filtered.

Part-of-speech Filtering. Research has provided varying results with respect to which parts of speech should be filtered using LDA [8, 29]. Against this background and taking into account that CVs contain a word distribution different from other types of documents (e.g., higher prevalence of nouns), part-of-speech filtering needs to be analyzed and adapted to obtain optimal results. In any case, nouns are not to be filtered as they transport essential information regarding competences.

Stemming & Lemmatization. Both stemming and lemmatization aim to decrease the number of considered terms by consolidating similar words. Stemming strives to truncate words to their stem, while lemmatization seeks to reduce words to their dictionary form. Stemming is seen as problematic for the application of LDA [27, 30, 31], for instance due to the danger that words with substantially different meaning are consolidated. Lemmatization, on the other hand, has mostly shown positive effects [29, 31]. However, CVs are structured differently than other types of documents, for example with respect to parts of speech; hence, the use of lemmatization should also be analyzed and adapted with respect to the database at hand to obtain optimal results.

Named Entity Recognition. Approaches for named entity recognition classify named entities in text into pre-defined categories (e.g., person names). The appearance of names in CVs is particular. CVs contain the name of the CV's author, possibly other person names (e.g., co-authors of publications), location names (referring to, e.g., company sites) and organization names. Location names may be useful in order to pinpoint expertise in certain areas such as the D-A-CH region. Organization names are of high relevance for many descriptions of competences (e.g., Microsoft Office). Person names, however, do not contribute to interpretable topics and should be filtered.

Overall, it has to be stated that – as it is usually the case in text mining – finding a very good pre-processing configuration for topic modeling of CVs is a non-trivial task. One has to experiment with different configurations to obtain optimal results with respect to the database at hand. Based on the discussion above, in the context of CVs, it seems particularly sensible to fix most steps but to experiment with n-gram creation, part-of-speech filtering and lemmatization.

3.4 Application of LDA

LDA requires as input two hyperparameters α and β as well as the total number of topics N. The shape of the CV-topic-distributions is determined by α [13]. When α is large, CVs are described by many topics and thus competences, whereas a small α leads to few topics per CV. Obviously, α should neither be too large (resulting in an unwieldy description of CVs which does not carve out the main competences) nor too small (resulting in only the most prominent competence being identified). The shape of the word-topic-distributions is controlled by β [13]. A large β implies that topics are widespread (i.e., competences are described broadly). A small β, in turn, leads to narrow topics and competences. In practice, α and β are often set to standard values (e.g., $1/N$) which have been shown to work well for a large range of application contexts [8]. Alternatively, an optimization can be performed [32].

If the number of topics N is too small, resulting topics may be general and widespread, representing a large variety of competences. For example, in a database containing CVs from IT experts, programming skills may constitute a single topic and not be differentiated further. As a result, topic distributions of CVs are not very meaningful: The competences of two persons portrayed by CVs with a similar topic distribution may still differ substantially. On the other hand, the more topics, the more challenging it is for humans to grasp all word-topic-distributions and to interpret CV-topic-distributions. Moreover, if N is too large, resulting topics may be very similar to each other. Thus, (almost) the same competences can be represented by multiple topics, leading to interpretation difficulties. The competences of two persons portrayed by CVs with a largely differing topic distribution may still be similar. To determine a favorable number of topics N, evaluation methods for topic models (cf. Section 2.1) can be used. Then, the resulting topics can be analyzed with regard to their human interpretability. In particular, it can be checked whether competences of interest are represented by topics or pre-processing configuration and LDA application need to be refined.

3.5 Interpretation & Use of Results

Once the topic model has been constructed, the actual knowledge discovery can begin. The topic model provides both word-topic-distributions and CV-topic-distributions.

The word-topic-distributions can be analyzed to obtain an understanding of the subjects generally present in the CVs. On a more fine-grained level, analyses of each topic – in particular, of the words with the highest probability in each topic – can be conducted to allow for their interpretation. Ideally, many of the topics clearly represent specific competences (e.g., web development). It is to be expected that also other topics representing, for instance, university or school career are contained in the model. In any case, as long as topics are interpretable, they should be labelled accordingly. Preferably, multiple persons label topics independently and compare their assessments afterwards.

The CV-topic-distributions offer a succinct description of each CV's topics. They allow to analyze CVs with respect to contained topics, in particular with respect to the competences of the portrayed person. This is especially helpful when topics have already been labelled. Then, the competences of a person portrayed by a CV can be assessed rapidly by observing the respective CV-topic-distribution and taking into account the labels associated to the most prevalent topics. Such an assessment is useful in HRM (e.g., for swift decision support in regard to the relevance of applicants for a job offer). Here, it is also important to note that using LDA, the topic distribution of a fresh CV can be determined quickly without re-running the whole model. Moreover, based on CV-topic-distributions, CVs may be categorized or tagged for future use. For example, all CVs which exhibit a proportion above 40% for a topic representing web development skills can be marked as relevant for future job offers in this area.

Furthermore, based on word-topic-distributions and CV-topic-distributions, techniques for post-processing LDA results can be applied. This includes in particular visualization approaches (e.g., [33, 34]) which assist analysts in gaining an overview and interpreting. For instance, LDAvis [34] provides clear illustrations of word-topic-distributions and offers to display terms particularly characteristic for a topic based on a relevance metric. This can be helpful to pinpoint rare but valuable competences occurring almost exclusively in a certain topic. If a specialist possessing this rare skill is required, the respective CV can then be retrieved quickly.

Obviously, there are further possibilities yet to be explored. We describe the following technique of a *topic-based search for CVs* as an exemplary idea. To facilitate this technique, in a first step, topics of interest and interpretable as competences are extracted from the LDA model and labelled. Based upon these topics, a query can be formulated which represents the desired competences in form of a search vector. For instance, if a Java developer with complementary competences in web development and web design is in demand, the emphasis may be 50% on Java programming, 30% on web development and 20% on web design. The search vector would thus include the values 0.5, 0.3 and 0.2 for topics representing Java programming, web development and web design respectively and 0 for all other topics. Then, the similarity between the search vector and the CV-topic-distributions of each CV can be calculated based on established similarity measures such as cosine similarity or Kullback-Leibler divergence for topics [17]. The most similar CVs can be manually screened and promising candidates can be contacted for recruiting. Such a topic-based search possesses clear advantages compared to a usual keyword search (e.g., also on platforms such as *LinkedIn* and *Indeed*), which can be illustrated by the example above:

1) The topic-based search allows to search for actual competences and not just for words which may or may not represent these competences reasonably well. For instance, a CV may not contain the term "web development" but the portrayed person may still report experience in JavaScript, HTML and PHP. The respective CV would be deemed irrelevant by a keyword search, whereas the topic-based search acknowledges the competence in web development.

2) In the topic-based search, it is possible to put emphasis on an aspect and the mere occurrence of a keyword is not enough for a CV to be relevant. For example, the specification that Java skills should make up for 50% of a CV's topic distribution means that CVs indeed need to contain a lot of Java-related terms to be assessed as relevant in the topic-based search. In contrast, a keyword search for terms such as Java is not very promising because a large number of CVs will claim at least some competence in Java.

3) The topic-based search allows to specify a weighting between different competences. In the example, Java programming is weighted with 50%, web development with 30% and web design with 20%. However, in a simple keyword search, each keyword would be treated as equally important. Weighting allows to more accurately identify the candidates which fit the requirements best (e.g., that Java development skills are most important and the other competences are complementary).

4 Application and Evaluation

In this section, we first describe how we exemplarily applied the procedure presented in Sections 3.1-3.5. In addition to this demonstration of practical applicability, we also evaluate the feasibility of the results within Section 4.5.

4.1 Acquisition of CVs

For our exemplary application, we decided to focus on CVs from IT experts for the following reasons. First, IT experts often possess diverse competences (e.g., programming languages, software, ...) which they report in their CVs. A topic modeling approach adapted to CVs should be able to identify these competences and categorize them into interpretable topics. Second, focusing on a single area provides a particular challenge for the procedure. CVs portraying persons with completely different competences are relatively easy to distinguish. However, a procedure that provides good results even when applied to CVs which are quite similar to each other – as in this case, CVs from IT experts – is of greater practical usefulness because more fine-grained distinctions can be made. Third, projects in companies often require IT experts with specific competences and thus, this application context is highly important.

To obtain CVs from IT experts, we used a web crawler (cf. Section 3.1, c)). This choice was made to allow the acquisition of CVs from private homepages, which many IT freelancers maintain. The web crawler was fed with search terms commonly used in IT (e.g., "Java") and including "CV". Based on these search terms and starting from the *Google* search, the web crawler stored approximately 27,000 PDFs.

4.2 General Pre-processing

To ready the collection for text pre-processing, we had to resolve the challenges (C1) and (C2) described in Section 3.2. PDFs were stored in a MongoDB database and converted to TXT for further analyses.

In order to address (C1) the diverse languages present in the collection, we performed automatic language identification and partitioned the collection accordingly. To this end, we used a Java open source tool [35]. On a manually inspected test sample of 100 documents, the approach did not produce any errors. We proceeded with German documents as those represented the largest proportion of the collection.

With regard to (C2), we observed that a quite large percentage of documents were not actually CVs. Thus, we manually classified documents into CVs and non-CVs to obtain a training dataset and built a classification model based on majority voting of the common classification methods logistic regression, support vector classification and random forests. The classification into CVs and non-CVs reached an accuracy of 95% on a test dataset of 1,291 documents and was applied to the remainder of the collection.

Overall, general pre-processing resulted in a database of 2,410 (presumed) CVs in German language which we used for further analyses. An exploratory data analysis was performed to obtain an overview. For instance, we determined the number of total (3,504,014) and unique (242,416) terms in the database and analyzed which terms occurred most frequently (all of them common stop words for German texts).

4.3 Text Pre-processing

The pre-processing routines with special characteristics in regard to CVs were set up as follows: In order to determine a comprehensive stop word list, we started by integrating established lists [36, 37]. Then, we modified this list by incorporating CV-specific stop words based on own reflections as well as an analysis of the 1,500 most frequent words in the database. The CV-specific stop words included in particular time specifications, legal forms of organizations and forms of address. Numbers were filtered as well. In contrast, terms such as "C" or "R" were removed from the list, as they represent ability in the respective programming languages in the given context.

To create n-grams, we used the NPMI-based method from the Python topic modeling library gensim [38]. 2-grams were created for terms with a NPMI value of at least 0.5 and a joint occurrence frequency of at least 100. Analogously, 3-grams were created from 2-grams (e.g., "Microsoft Visual" and "Visual Studio" were combined to "Microsoft Visual Studio") and so on. In this way, many n-grams were created, the most frequently used ones being "SQL Server", "SAP R3" and "MS Office".

Part-of-speech tagging was realized by the aggregated results of two taggers. We used TreeTagger [39], a tagger based on a probabilistic Markov model pre-trained for the German language, and the Natural Language Toolkit [40] tagger which was trained with the TIGER corpus [41]. On a manually inspected test sample of 400 words, tagging in this way exhibited an accuracy of 96%. TreeTagger was also used for lemmatization.

For named entity recognition, the Stanford NER Tagger [42] was used. Its results were then refined by publicly available lists of first names [43, 44] and a phone book for surnames, achieving a true positive rate of 97% in regard to filtered person names.

As suggested in Section 3.3, we fixed many of the pre-processing routines but let others vary and experimented in order to achieve optimal results. To be more precise, we always – and in this order – removed formatting tags and special characters, tokenized and lowercased the CVs and removed stop words and words occurring only in few documents. We experimented with n-gram-creation (yes/no), lemmatization (yes/no), part-of-speech filtering (possibly filtering adjectives and/or verbs and/or adverbs) as well as the threshold for words to occur in too few documents (50/40/30/20/10). Overall, this resulted in 160 pre-processing configurations.

For each configuration, LDA models were generated using the gensim libary [38] and each number of topics $N \in \{25,50,75,100\}$, following [32]. Convergence was tested as suggested in [15]. Subsequently, the generated LDA models were evaluated with respect to semantic coherence (cf. Section 2.1). We determined the configuration leading to the highest value of semantic coherence and used it for optimizing the LDA application (cf. Section 4.4). Thereafter, to verify the results, we re-ran the evaluation of all configurations with the optimized LDA model. The configuration leading to the highest value of semantic coherence (NPMI: 0.161) did not consider n-gram-creation and lemmatization, filtered adjectives as well as verbs and adverbs, and filtered all words occurring in less than 40 documents.

The results in regard to n-gram-creation and lemmatization may surprise at first. However, they are in line with previous research in other application contexts suggesting that LDA derives required semantic relations itself and stronger pre-processing reduces topic model quality [30]. Similarly, filtering all parts-of-speech except nouns has also already been shown to provide strong results [29] and is expected to be promising for CVs, with competences usually being described by nouns.

4.4 Application of LDA

The hyperparameters α and β were optimized similar to the pre-processing configuration, again following [32]. To determine the number of topics N, we generated LDA models for each $N \in \{2,4,6,...,100\}$ based on the optimal pre-processing configuration. We then analyzed semantic coherence for each N. This led to choosing

N=42, a number of topics manageable for humans. Thus, the respective topic model was examined further with respect to interpretability and use of results.

4.5 Interpretation & Use of Results

The interpretation of the topic model was conducted separately by two human coders to account for human subjectivity. Consolidating the interpretation only required to settle minor wording differences. The word-topic-distributions of each topic were analyzed to obtain an overview of the topic model. We observed that topics could generally be categorized into three groups: Group A, the largest group, contained topics describing specific IT-related competences and consisted of 23 topics. The four topics in Group B described competences concerning business & management. The remaining 15 topics in Group C were related to university or school career and, due to our focus on competences, not considered for further analysis. Some topics are shown exemplarily in Table 1. Thereby, a topic is represented by its seven words with highest probability in decreasing order and translations for German words are provided in square brackets. Please note that many English words are used frequently in German CVs, explaining their occurrence in topics.

Table 1. Exemplary topics from each of the groups A, B, C

ID	*Gr.*	*Topic (most probable words)*
1	A	java eclipse entwicklung[development] xml spring j2ee oracle
2	A	linux server administration unix system perl security
3	A	design adobe konzeption[conception] 3d web programmierung[programming] photoshop
4	A	entwicklung[development] web javascript php mysql css html
5	A	windows server ms microsoft support software office
6	A	c r analysis time solution network networks
7	B	management projekt[project] einführung[launch] analyse[analysis] projektmanagement[project management] business durchführung[execution]
8	C	university research school international european science german

Overall, the analysis yielded that most of the topics in Groups A and B are fine-grained topics clearly representing specific competences. More precisely, they do not describe competences rather general for CVs from IT experts such as programming skills, but more distinguishing competences such as programming skills in Java (cf. Topic 1). Employing the relevance metric of LDAvis [34], it was possible to differentiate topics even further and carve out competences highly characteristic for a topic. Usually, this concerned closely associated special frameworks, software or technologies. For instance, the most relevant words of Topic 1 (describing programming skills in Java) then were: jaxb, j2se, jpa, jax, ejb, hibernate, jms. All of them are Java-specific and occurred almost exclusively in Topic 1.

The results support that the topics possess a clear interpretation with respect to describing competences and are thus useful for HRM. Furthermore, based on these results, a topic-based search seemed promising. We labelled 21 of the 27 topics in Groups A and B with respect to the competence described (the remaining 6 topics were judged to not be as clearly interpretable and left out). Besides their use to rapidly assess the competences represented in a CV for HRM (cf. Section 3.5), the labels facilitated the topic-based search. Our prototypical implementation allows the specification of search queries as vectors containing the desired weight for each topic. Similarities between the search vector and the CV-topic-distributions of each CV are calculated based on Kullback-Leibler divergence [17]. The most similar CVs are shown together with their topic distribution. Figure 2 illustrates the GUI of the prototypical implementation with the search query from Section 3.5 and the first search result (CV #545). Clicking on a search result opens the respective CV.

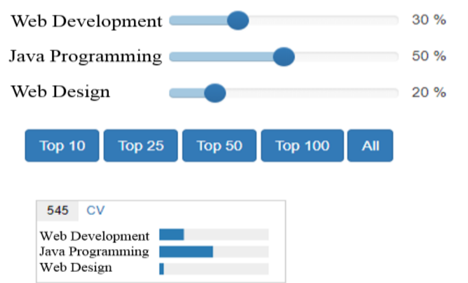


Figure 2. Topic-based search for CVs

CV #545 portrayed a senior Java developer with skills in a large number of Java-related technologies (Spring, Struts, JUnit, JEE, ...). The CV also claimed a lot of work expertise in web development such as the programming of web frontends using HTML and CSS. To a lesser extent, competences in web design and respective software (e.g., Photoshop, Gimp) were reported as well. Thus, the CV fit the job offer represented by the search query exceptionally well. The analysis of the other top 10 CVs which were determined to be the best match for the search query yielded similar results.

For comparison, we also performed a keyword search using the search term *java AND "web development" AND "web design"*. Here, we observed all 3 advantages of a topic-based search outlined in Section 3.5: With respect to 1), the keyword search only yielded 4 results because few CVs actually followed the exact wording dictated by the search term. In particular, many suitable CVs such as the ones found by our topic-based search were neglected by the keyword search. Regarding 2), the problem of merely focusing on keywords became obvious as the CV of a manager who once had conducted a Java project was included in the 4 results of the keyword search, but did in fact not fit the job offer. Concerning 3), the lack of weighting showed when the CV of a web developer specializing in PHP, HTML and JavaScript with basic Java abilities was assessed as relevant. Overall, none of the results of the keyword search fit the job offer well. Our topic-based search thus produced clearly superior results in this setting.

We further specified 8 more search queries and analyzed the respective CVs suggested by the topic-based search. In each case, the competences reported in the top

CVs coincided with the competences called for by the search query. To conclude, the topic-based search worked very well in this application and seems fit to provide helpful decision support for HRM.

5 Conclusion, Practical Implications and Directions for Future Work

In this paper, a topic modeling procedure consisting of five steps with the aim of discovering knowledge from CVs has been presented. CV-specific characteristics are considered in each step. An exemplary application to CVs from IT experts suggests that clearly interpretable topics describing fine-grained competences (e.g., Java programming, web design) can be discovered. This information can be used to rapidly assess the contents of a CV, categorize CVs and identify promising candidates for job offers, thus providing decision support in HRM.

The presented procedure allows for proactive recruiting. It can, for instance, be applied in HRM similar to how professional social networks are currently used in the recruitment process to rapidly source candidates before subsequent steps such as job interviews are conducted. However, it is not restricted to members of these networks as the analyzed CVs may stem from any origin. Additionally, the presented topic-based search possesses advantages compared to a keyword search on these platforms. Moreover, the CV-topic-distributions in conjunction with labels can be used to categorize and tag CVs for future use. In this way, companies can construct and steadily extend a database of interesting CVs. Another promising idea for companies is to also include CVs of own employees to promote internal recruiting. In any case, HRM and IT departments need to cooperate as skills from both areas are required to achieve a successful application of the procedure.

While the paper at hand offers a detailed description of a procedure for knowledge discovery from CVs, there are also limitations which provide directions for further research. First, an application to CVs from a different context should be conducted to validate feasibility. Second, a topic-based search technique has been presented and evaluated in an exemplary setting. However, it should be further assessed, for instance by a more detailed comparison to alternatives (e.g., with the help of a HRM expert) and an application to real job offers from a company. Finally, CV-specific visualization approaches should be developed, allowing for an easier overview and use of the results of the procedure. They should be included in a tool facilitating and partly automating the five steps of the procedure to further its practical use.

References

1. Breaugh, J.A.: Employee recruitment: Current knowledge and important areas for future research. Human Resource Management Review 18, 103–118 (2008)
2. Hendry, C.: Human resource management. Routledge (2012)
3. Allden, N., Harris, L.: Building a positive candidate experience: towards a networked model of e-recruitment. Journal of Business Strategy 34, 36–47 (2013)

4. Gao, L., Eldin, N.: Employers' expectations: A probabilistic text mining model. Procedia Engineering 85, 175–182 (2014)
5. Abel, F., Deldjoo, Y., Elahi, M., Kohlsdorf, D.: Recsys challenge 2017: Offline and online evaluation. In: Proceedings of the 11th ACM RecSys, pp. 372–373 (2017)
6. LinkedIn: About LinkedIn, https://about.linkedin.com/
7. Coffey, A., Atkinson, P.: Making sense of qualitative data: complementary research strategies. Sage Publications, Inc (1996)
8. Debortoli, S., Müller, O., Junglas, I., vom Brocke, J.: Text mining for information systems researchers: an annotated topic modeling tutorial. CAIS 39, 111–135 (2016)
9. Strohmeier, S., Piazza, F.: Domain driven data mining in human resource management: A review of current research. Expert Systems with Applications 40, 2410–2420 (2013)
10. Gupta, V., Lehal, G.S.: A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence 1, 60–76 (2009)
11. Blei, D.M.: Probabilistic topic models. Communications of the ACM 55, 77–84 (2012)
12. Gorbacheva, E., Stein, A., Schmiedel, T., Müller, O.: The role of gender in business process management competence supply. BISE 58, 213–231 (2016)
13. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (2003)
14. Belford, M., Mac Namee, B., Greene, D.: Stability of topic modeling via matrix factorization. Expert Systems with Applications 91, 159–169 (2018)
15. Hoffman, M., Bach, F.R., Blei, D.M.: Online learning for latent dirichlet allocation. In: NIPS 23, pp. 856–864 (2010)
16. Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: NIPS 22, pp. 288–296 (2009)
17. Koltcov, S., Koltsova, O., Nikolenko, S.: Latent dirichlet allocation: stability and applications to studies of user-generated content. In: Proceedings of the 6th ACM WebSci, pp. 161–165 (2014)
18. Aletras, N., Stevenson, M.: Evaluating topic coherence using distributional semantics. In: Proceedings of the 10th IWCS, pp. 13–22 (2013)
19. Lau, J.H., Newman, D., Baldwin, T.: Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In: Proceedings of the 14th EACL, pp. 530–539 (2014)
20. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: Proceedings of the 8th NAACL, pp. 100–108 (2010)
21. Röder, M., Both, A., Hinneburg, A.: Exploring the space of topic coherence measures. In: Proceedings of the 8th ACM WSDM, pp. 399–408 (2015)
22. Fang, D., Yang, H., Gao, B., Li, X.: Discovering research topics from library electronic references using latent Dirichlet allocation. Library Hi Tech (2018)
23. Guo, Y., Barnes, S.J., Jia, Q.: Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tourism Management 59, 467–483 (2017)
24. Dumont, T., Fettke, P., Loos, P.: Towards multi-dimensional Clustering of Business Process Models using Latent Dirichlet Allocation. Tagungsband zur Multikonferenz Wirtschaftsinformatik, 69–80 (2016)
25. Shuyo, N.: Language detection library, https://www.slideshare.net/shuyo/language-detection-library-for-java
26. Jauhiainen, T., Lui, M., Zampieri, M., Baldwin, T., Lindén, K.: Automatic language Identification in texts: A survey. arXiv preprint arXiv:1804.08186 (2018)

27. Boyd-Graber, J., Mimno, D., Newman, D.: Care and feeding of topic models: Problems, diagnostics, and improvements. Handbook of mixed membership models and their applications, 225–255 (2014)
28. Weiss, S.M., Indurkhya, N., Zhang, T., Damerau, F.: Text mining: predictive methods for analyzing unstructured information. Springer Science & Business Media (2010)
29. Martin, F., Johnson, M.: More efficient topic modelling through a noun only approach. In: Proceedings of the ALTA 2015, pp. 111–115 (2015)
30. Schofield, A., Mimno, D.: Comparing apples to apple: The effects of stemmers on topic models. Transactions of the ACL 4, 287–300 (2016)
31. Spies, M.: Topic modelling with morphologically analyzed vocabularies. Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics, Informatics and mechanics 9, 1–18 (2017)
32. Wallach, H.M., Mimno, D.M., McCallum, A.: Rethinking LDA: Why priors matter. In: NIPS 22, pp. 1973–1981 (2009)
33. Chaney, A.J.-B., Blei, D.M.: Visualizing Topic Models. In: ICWSM (2012)
34. Sievert, C., Shirley, K.: LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp. 63–70 (2014)
35. Shuyo, N.: Language detection, https://github.com/shuyo/language-detection
36. Salton, G. and Buckley, C.: Stop word list 2, http://www.lextek.com/manuals/onix/stopwords2.html
37. Götze, M. and Geyer, S.: German stopwords, https://github.com/solariz/german_stopwords/blob/master/german_stopwords_full.txt
38. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50 (2010)
39. Schmid, H.: TreeTagger - a part-of-speech tagger for many languages, http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
40. NLTK Project: Natural Language Toolkit, https://www.nltk.org/
41. Institut für maschinelle Sprachverarbeitung: TIGER corpus, http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html
42. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd ACL, pp. 363–370 (2005)
43. Michael, J.: Anredebestimmung anhand des Vornamens, https://www.heise.de/ct/ftp/07/17/182/
44. Kolb, P.: Liste mit Vornamen, http://www.ling.uni-potsdam.de/~kolb/Vornamen.txt

Online Product Descriptions – Boost for your Sales?

Tristan Wimmer[1], Michael Scholz[1]

[1] University of Passau, School of Business, Economics and Information Systems, Passau, Germany
{michael.scholz,tristan.wimmer}@uni-passau.de

Abstract. Product descriptions are a source of information online consumers can use to reduce product uncertainty. Recent research provides evidence that consumers favor using information from other consumers, such as customer reviews, over retailer or manufacturer provided information, such as product descriptions. We complement this research and show that the presence of product descriptions significantly influences products' sales and that this influence decreases with an increasing number of customer reviews. We furthermore demonstrate that a product description's information amount positively affects a product's sales. The number of customer reviews available for a product also moderates the effect of the information amount of a product description on sales.

Keywords: Product Description, Product Uncertainty, E-Commerce, Customer Reviews

1 Introduction

Online consumers face a barrier in physical experience of products. While consumers in offline markets can touch the product of their choice, online consumers hardly can evaluate a product's physical characteristics prior to purchase. Consumers in online as well as offline markets typically perceive uncertainty in purchase decision processes [1–3].

Pavlou et al. [4] investigated reasons for perceived product uncertainty by considering the relationship between sellers and consumers as a principal-agent problem. Amongst others, they identified information asymmetry as one of the most important determinants of perceived uncertainty. This information asymmetry refers to the seller or the product and finally leads to seller uncertainty or product uncertainty [4, 5]. Seller uncertainty is defined as consumer's difficulty to predict a seller's behavior in the future whereas product uncertainty refers to the difficulty to evaluate a product's quality prior to purchase [2, 6].

Product descriptions often are the only information source for consumers. Customer reviews and expert tests especially are not available for new products. In contrast to customer reviews, sellers furthermore can control the content of product descriptions. Thus, product descriptions likely will influence sales. However, product descriptions and their impact on sales have been sparsely addressed in recent research. We focus on

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the effect of product descriptions on the reduction of product uncertainty. We follow Pavlou et al. [4] and assume that a lower product uncertainty leads to higher sales and analyze the impact of product descriptions on sales. The research question investigated in this paper therefore is as follows:

RQ: What is the effect of product descriptions on product sales?

Our empirical investigation shows that products with a product description generate higher sales than products without a description. We provide evidence that the information amount of product descriptions is positively correlated with products' sales.

With this paper, we contribute to recent research by (1) examining the effect of the presence and the information amount of product descriptions on sales, (2) distinguishing between the effect of product descriptions written by a retailer and product descriptions written by a manufacturer, (3) investigating the interaction effect of product descriptions and customer reviews on sales.

The remainder of this paper is organized as follows. In the next section, we review related research. We thereafter discuss the theoretical background on product uncertainty and its reduction. Afterwards, we describe our research model followed by a description of our empirical evaluation. We then present the result of the empirical evaluation and of some robustness checks. We conclude this paper with a summary of our results and a discussion of the implications for researchers and practitioners.

2 Related Research

Consumers in offline markets can inspect a product's physical characteristics and can get personal advice from the seller. Consumers in online stores, such as Amazon or Staples, predominantly have two sources of information to learn about a product's characteristics: customer reviews and product descriptions [7]. Consumers typically have a higher trust in information from other consumers than from a producer or a marketplace [8]. Recent research thus has intensively investigated the effect of customer reviews, as one source of information, on consumers' purchase decision processes and retailers' sales [9–11]. These studies demonstrate that customer reviews affect purchase decisions and ultimately sales to a large extent. Wang et al. [12] investigated customer reviews embedded in product descriptions and found, that embedded customer reviews rise sales.

The more reviews are available, the lower is a consumer's perceived product uncertainty [13, 14]. But if there are no or only a few customer reviews available, consumers tend to consult other information sources (e.g., Question and Answer (Q&A) technologies or product descriptions). Banerjee et al. [15] found evidence, that question and answer technologies can reduce a consumer's uncertainty, whether the product fits her needs, affecting the review quality. However, implementing such Q&A technologies is not costless and thus, most market places do not provide such technologies which in turn forces consumers to contact other information resources such as product descriptions. Dimoka et al. [2] found evidence that product descriptions reduce product uncertainty most significantly among several information sources, such as product descriptions, product inspections, history reports and product warranties. Further, Detlor et al. [16] found evidence, that product descriptions are important in pre-purchase online information seeking. Hence, product descriptions are another important source for reducing product uncertainty with two major advantages compared to customer reviews: First, product descriptions are also available in the absence of customer reviews, because they are provided by the seller. Thus, product descriptions are often the only source of information consumers can digest to learn about a product's characteristics prior to purchase. And second, sellers can control product descriptions and hence the extent to which they reduce product uncertainty.

We discuss the relationship between product uncertainty and sales and how to reduce product uncertainty in order to improve sales in the next section.

3 Theoretical Background

3.1 Product Uncertainty and Sales

Product uncertainty is defined as consumers' difficulty to evaluate a product's characteristics prior to purchase [17]. The higher the variance of product characteristics the higher is the perceived product uncertainty [6]. A consumer who wants to purchase a new electric toothbrush might feel uncertain about whether a particular toothbrush is controllable via a mobile application. If there is no toothbrush with mobile application support available, there is no variance for this characteristic. Thus, in the case that none of the toothbrushes provides mobile application support there is no product uncertainty. As soon as there are toothbrushes with and without mobile application support available, consumers perceive product uncertainty to a particular extent about the quality of that mobile application support.

Product uncertainty has negative implications for both, sellers and consumers. Recent research has demonstrated that the higher the product uncertainty, the lower is the price premium that can be charged for a particular product [2]. Product uncertainty furthermore negatively affects sales [4] and the number of product returns [6]. Consumers' transaction costs increase with the perceived product uncertainty because consumers need to invest search costs in order to reduce product uncertainty [18]. Furthermore, a rising product uncertainty decreases consumers' purchase intention significantly [4]. Thus, sellers and consumers have a keen interest in reducing product uncertainty. Therefore, sellers and consumers provide information in form of product descriptions or customer reviews to reduce product uncertainty in online stores.

Recent research provided evidence that sales significantly rise when consumers' uncertainty in product quality decreases. Given the positive relationship of low product uncertainty and sales [4], sellers seek to reduce product uncertainty in order to improve their sales. Besides motivating consumers to submit product reviews, sellers also can provide product descriptions that discuss important product features. In the next section, we discuss how sellers can reduce product uncertainty in order to improve their sales in more detail.

3.2 Reduction of Product Uncertainty in Online Stores for Improving Sales

Online consumers want to learn about a product's characteristics before purchase in order to reduce product uncertainty. In offline stores, consumers can reduce product uncertainty by inspecting a product itself and requesting individual advice from the seller. Inspecting products prior to purchase is typically not possible in online stores. Thus, online consumers gather information about a product's characteristics from different information sources, such as product descriptions and customer reviews [19, 20].

Customer reviews are peer-generated product evaluations that typically consist of a product rating and an optional textual description of the experiences with the product [21, 22]. Recent research has shown that the existence of customer reviews reduces product uncertainty and thereby improves consumers' purchase probability [20]. Customer reviews might vary significantly in the product characteristics they discuss. The probability that a consumer can learn about some certain characteristic increases with the number of provided customer reviews [23]. Mudambi and Schuff [21] furthermore find evidence that the length of the textual description of a customer review is positively correlated with the perceived helpfulness of the review. Longer reviews presumably discuss more product characteristics. Archak et al. [24] show that reviews discussing more product characteristics are more influential on consumers' purchase decisions. Similarly, Scholz and Dorner [22] find support that reviews with a higher information amount are more helpful for consumers. Reviews with a higher helpfulness are more likely to reduce product uncertainty [21]. Helpfulness has been furthermore shown to positively affect a retailer's sales [9]. In summary, product reviews are an important source of information. They influence consumers' decision-making processes by reducing information asymmetries and thus by reducing product uncertainty.

In contrast to customer reviews, the effect of product descriptions on reducing consumers' product uncertainty has been sparsely analyzed in existing research. Dimoka et al. [2] investigate the influence of product descriptions on product uncertainty in online car auctions. They provide evidence that the influence of product descriptions is nearly twice as much as that of third-party assurances, such as (car) inspections, history reports or product warranties. Ghose and Han [7] find evidence that the length of product descriptions is positively correlated with sales.

In summary, existing research demonstrates that customer reviews and product descriptions contribute to reduce product uncertainty. A positive impact of customer reviews on sales has also been found in several studies [25–28] . An investigation of the effect of the existence of product descriptions as well as the interaction of product descriptions and customer reviews on sales is missing so far.

4 Research Model

Recent research has illustrated that product uncertainty significantly influences sales [4]. Factors positively influencing product uncertainty are hence likely to negatively influence sales. In the following, we will use sales as proxy for product uncertainty because product sales are easily observable.

Product descriptions have been found to be one source for reducing product uncertainty [2]. If no or not enough information is available for reducing product uncertainty, consumers refrain to buy, especially high-priced products [29]. However, not all products do have a product description being available in an online store. Even the availability of product descriptions might help consumers to learn about product characteristics and thereby reduce product uncertainty and improve sales. We thus hypothesize a positive effect between the availability of a product description and the product's sales.

H1a: Products with a product description generate on average more sales than products without a product description.

The information amount transported in a product description significantly varies across the products. Descriptions discussing more product characteristics are more likely to help consumers reducing product uncertainty. Recent research has provided evidence that customer reviews, as another source of information for consumers, are perceived as more helpful if they discuss more product characteristics that are not widely discussed in other customer reviews [22, 30]. Such an effect has been shown for app descriptions [7]. We expect a similar effect also for product descriptions and thus hypothesize that a product's sales are increasing in the product description's information amount.

H1b: The higher the information amount of a product's description, the higher are its sales on average.

H1a can be seen as a special case of hypothesis H1b. A product description that is not available is equivalent to a product description with an information amount of zero.

Recent research provides ample evidence that the number of customer reviews available for a particular product positively influences this product's sales [14, 28]. Consumers prefer customer reviews over product descriptions because the retailer or manufacturer generates the latter [8]. Product descriptions communicate a positive picture about a product whereas customer reviews also point to a product's drawbacks. However, product descriptions exist, in contrast to customer reviews, even if no consumer has bought or evaluated the product. The more reviews are available, the higher is the probability that a particular consumer will find enough information to reduce product uncertainty and the higher will be the probability to purchase a product. Recent research has demonstrated that the number of reviews positively affects sales [9, 31]. We therefore propose a decreasing impact of a product's description on sales when there is an increasing number of customer reviews available.

H2: The influence of the availability of a product description on a product's sales is moderated by the number of reviews. The more reviews a product has, the lower is the effect of the availability of a description on sales.

Similarly, the impact of a description's information amount might be also moderated by the number of reviews available for a product.

H3: The influence of a description's information amount on sales is moderated by the number of reviews. The more reviews a product has, the lower is the effect of its descriptions' information amount on sales.

Existing research has demonstrated a positive effect of the number of reviews on sales [28]. More customer reviews lead to more sales and hence more customers that are prepared to write further customer reviews. We follow these findings and also propose a positive effect of the number of customer reviews on sales.

H4: The more customer reviews are available for a product the higher are this product's sales.

5 Empirical Evaluation

In order to prove our hypotheses, we collected data for three product categories from Amazon.com. We collected data about 84 backpacks, 131 pencils and 136 electric toothbrushes for a period of 39 days. For each product of each category and day, we gathered the price, the average rating and the number of reviews. Amazon provides typically two kinds of product descriptions: descriptions generated by the manufacturer and descriptions generated by the retailer (Amazon.com). Thus, we collected the product description provided by the manufacturer and the product description provided by Amazon.com for each product. Figure 2 represents an exemplary product description of an electric toothbrush.

The price range for each product category is rather large as shown in Table 1. On the one hand, each category has at least one product which doesn't have any customer reviews (ratings) and on the other hand, each category includes products with a lot of customer reviews (e.g., there are more than 1,000 reviews available for 10 of the electric toothbrushes in our dataset). An Amazon product description is available for 90% of the backpacks, 96% of the pencils and 96% of the electronic toothbrushes in our dataset. Manufacturer product descriptions are significantly less often available than Amazon product descriptions ($p < 0.001$). Only 23% of the collected backpacks, 3% of the pencils and 16% of the electronic toothbrushes provide a manufacturer product description. The mean lengths of Amazon and manufacturer product description differ. Amazon product descriptions furthermore consists of significantly more words than manufacturer descriptions ($p < 0.001$). Amazon descriptions on average are 91.76 words long whereas manufacturer descriptions only consist of 48.13 words. We will use product sales ranks as proxy for sales because sales figures are not available on Amazon.com. Schnapp and Allwine [32], Chevalier and Goolsbee [33], as well as Chevalier and Mayzlin [26] found a linear relationship between ln(sales) and ln(rank). We thus use ln(rank) as dependent variable in order to allow a more straightforward interpretation of the estimated effects on sales. The lower the sales rank of a product, the more instances of this product have been sold compared to other products in the same category. The mean sales rank for the backpacks in our dataset is 3911.85, that for pencils is 244.24 and the toothbrushes in our dataset or on average on rank 307.18. For each product category, we collected top sellers (i.e., products with a sales rank of

1) as well as well as niche products (i.e., products with a rather high category-specific sales rank).

Figure 2. Exemplary Manufacturer and Amazon Product Description of an Electronic Toothbrush

A summary of the collected data variables is provided in Table 1 and Table 2.

Table 1. Variables Overview

Variable	*Symbol*	*Description*
Price	p	Price of the product
Average Rating	avr	Average rating based on customer reviews
Number of Reviews	rev	Number of customer reviews
Amazon Product Description	apd	Availability of an Amazon product description
Manufacturer Product Description	mpd	Availability of a manufacturer product description
Sales Rank	rank	Sales rank of the product
Day	d	Day at which the data have been collected
Length Amazon Product Description	la	Number of words in the Amazon product description
Length manufacturer Product Description	lm	Number of words in the manufacturer product description

Table 2. Descriptive Statistics of the Variables in the Dataset

Variable	*p*	*avr*	*rev*	*apd*	*mpd*	*rank*	*la*	*lm*
Backpacks (min)	7.99	-	0	0	0	10	0	0
Backpacks (max)	447.5	5	2822	1	1	72690	307	377
Backpacks (mean)	65.82	4.28	388.97	0.91	0.23	3911.85	91.76	48.13
Pencils(min)	1.99	-	0	0	0	1	0	0
Pencils(max)	56.39	5	2588	1	1	1935	322	404

Pencils(mean)	11.15	3.9	91.22	0.96	0.03	244.24	64.43	7.29
Toothbrushes (min)	3.8	-	0	0	0	1	0	0
Tooth-brushes (max)	199.99	5	6233	1	1	1466	510	1018
Tooth-brushes (mean)	44.55	3.95	316.47	0.95	0.16	307.18	145	93

6 Analysis and Results

6.1 Effect of Product Descriptions

In H1a, we hypothesized that products with an available product description generate on average more sales than products without a product description. According to H2, the number of customer reviews available for a product moderates this effect. H4 proposes a positive effect of the number of customer reviews on sales. In order to test H1a, H2 and H4, we estimate the effect of the presence of a product description binary-coded by the variables apd (Amazon product description) and mpd (manufacturer product description) on product i's sales rank rank at day d. We use price p, average rating avr and number of reviews rev as control variables, because they have been found to significantly affect sales ranks [11]. More specifically, we estimate the following model with fixed product and time effects for investigating H1a, H2 and H4:

$$\begin{aligned}\log\left(rank_{i,d}\right) = {} & \beta_1 \log(p_{i,d}) + \beta_2 avr_{i,d} + \beta_3 \log(rev_{i,d} + 1) + \beta_4 apd_{i,d} + \beta_5 mpd_{i,d} \\ & + \beta_6 \log(rev_{i,d} + 1) apd_{i,d} + \beta_7 \log(rev_{i,d} + 1) mpd_{i,d} \\ & + \beta_8 category + \gamma_1 d + \gamma_2 i + \varepsilon_{i,d}\end{aligned}$$

We use the logarithm of product i's price to model a diminishing effect of price on sales rank. A price difference between 200 and 210 Euros might be less relevant for a consumer than a price difference between 20 and 30 Euros. We also use the logarithm of each product's number of reviews to model a diminishing effect of this variable on the sales rank. Furthermore, we use the logarithm of each product's sales rank as dependent variable in order to model diminishing perceived differences between sales ranks [34]. Variance inflation factors of less than 2 indicate the absence of multicollinearity in our model. Based on a Durbin-Watson test, we did not find an indication for autocorrelated residuals (D = 1.9998, p = 0.497). The results of our regression model with robust standard errors are depicted in Table 3.

Table 3. Effect of Product Descriptions on log(Sales Rank)

Variable	*Estimate*	*Std. Error*	*t-Value*	*p-Value*
log(Price)	0.361	0.030	11.950	< 0.001
Average Rating	0.062	0.016	3.922	< 0.001
log(Number of Reviews+1)	-0.634	0.055	-11.506	< 0.001
Amazon Product Description	-0.778	0.159	-4.898	< 0.001
Manufacturer Product Description	-0.606	0.221	-2.742	0.006

log(Number of Reviews + 1) x Amazon Product Description	0.178	0.051	3.484	< 0.001
log(Number of Reviews + 1) x Manufacturer Product Description	0.083	0.028	2.926	0.003
Pencils	-3.163	0.106	-29.947	< 0.001
Toothbrushes	-2.496	0.134	-18.658	< 0.001
Adj. R^2 (full model)		0.960		
Adj. R^2 (projected model)		0.341		

As expected, Table 3 also shows that a lower price and a higher number of customer reviews result in a better sales rank. A somewhat counterintuitive result that emerges from Table 3 is that a better (higher) average rating increases the sales rank. One would expect that a better rating would decrease the sales rank. However, previous research has shown that there exists a negative relation between sales and average rating for some product categories [35]. Customer ratings typically follow a J-shape distribution with most ratings being very positive (i.e., they are 5-star ratings) [36]. Consumers might fear manipulated 5-star ratings and rather trust products being characterized not only by 5-star ratings, which ultimately might lead to a positive observed relation between sales rank and a product's average rating.

The results in Table 3 indicate that both types of product descriptions – Amazon and manufacturer descriptions – significantly influence a product's sales rank. The existence of an Amazon product description on average decreases the sales rank by 2.18 ranks whereas the existence of a manufacturer product description decreases 1.82 ranks. Amazon descriptions hence have a higher impact on a product's sales rank. Manufacturers should consequently create descriptions for their products in order to improve their sales. Our hypothesis H1, that the presence of a product description has a significantly positive influence on sales is therefore supported by our data.

In H2, we proposed that the effect of product descriptions on sales will diminish with an increasing number of customer reviews. Positive estimates of the interaction effects between the number of reviews and the availability of both, Amazon product descriptions and manufacturer product descriptions, show support for H2.

The negative estimate for the number of reviews indicates that more customer reviews lead to a better sales rank and hence more sales. This is in support of hypothesis H4.

Our results show a high importance of product descriptions for improving a product's sales. We will analyze the impact of the information amount of Amazon and manufacturer product descriptions in the next section.

6.2 Effect of Product Descriptions' Information Amount

In H1b, we hypothesized that a higher information amount in a product description will lead to a better sales rank. To test H1b, we use the number of words as a proxy for a description's information amount. H3 hypothesizes an interaction effect between the number of reviews and the information amount of product descriptions. We again test hypothesis H4 with this analysis. We use the following model to prove H1b, H3 and H4.

$$\log(rank_{i,d}) = \beta_1 \log(p_{i,d}) + \beta_2 avr_{i,d} + \beta_3 \log(rev_{i,d} + 1) + \beta_4 \log(la_{i,d} + 1) + \beta_5 \log(lm_{i,d} + 1) + \beta_6 \log(rev_{i,d} + 1) \log(la_{i,d} + 1) + \beta_7 \log(rev_{i,d} + 1) \log(lm_{i,d} + 1) + \beta_8 category + \gamma_1 d + \gamma_2 i + \varepsilon_{i,d}$$

With variance inflation factors of less than 2.1, we can assume that our model is not subject to multicollinearity. A Durbin-Watson test also indicated that our model is not subject to an autocorrelation problem (D = 1.9998, p = 0.496). The results in Table 4 again show that a lower price and a higher number of customer reviews result in a better sales rank. Table 4 also shows that the information amount of Amazon product descriptions has a significant positive influence on a product's sales rank (the higher the information amount, the lower is the sales rank). The effect of manufacturer descriptions is in the same direction but only weakly significant. Hypothesis H1b is hence supported partially.

H3 proposes a diminishing effect of the information amount of product descriptions on sales when the number of customer reviews increases. As shown in Table 4 and Figure 3, we found such a diminishing effect for both Amazon and manufacturer product descriptions. The interaction effect of the number of customer reviews and the information amount of manufacturer descriptions is, however, only weakly significant. The more customer reviews are available the higher is the estimate for the information amount indicating that a higher number of available customer reviews leads to a diminishing effect of the information amount on sales. Hypothesis H3 is therefore partially supported by our data.

Table 4. Effect of Product Descriptions' Information Amount on log(Sales Rank)

Variable	*Estimate*	*Std. Err.*	*t-Value*	*p-Value*
log(Price)	0.361	0.030	12.097	< 0.001
Average Rating	0.084	0.016	5.252	< 0.001
log(Number of Reviews + 1)	-0.744	0.051	-14.574	< 0.001
log(Length Amazon Description+1)	-0.209	0.030	-7.078	< 0.001
log(Length Manufacturer Description+1)	-0.088	0.048	-1.829	0.068
log(Number of Reviews + 1) x log(Length Amazon Description+1)	0.074	0.011	6.919	< 0.001
log(Number of Reviews + 1) x log(Length Manufacturer Description+1)	0.010	0.006	1.673	0.094
Pencils	-2.746	0.129	-21.218	< 0.001

Toothbrushes	-2.050	0.146	-14.043	< 0.001
Adj. R^2 (full model)			0.960	
Adj. R^2 (projected model)			0.344	

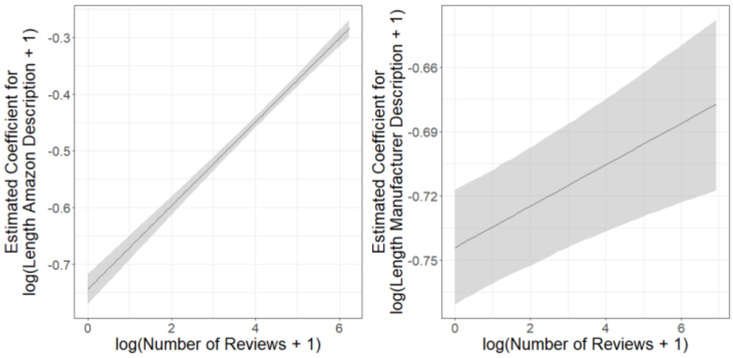


Figure 3. Interaction Effect Between the Availability of Product Descriptions and the Number of Reviews

We again found support for hypothesis H4. A higher number of customer reviews lowers the sales rank which is equivalent to an improvement in sales.

7 Robustness Check

We test the robustness of our results by investigating the effects of product descriptions on quantiles rather than average values of sales ranks. We found a very high variance of 13.075.579 for the sales ranks (2.547 for log(sales ranks)) in our data set indicating that our data set encompasses blockbusters as well as niche products. Thus, we analyze the effects of product descriptions on quantiles of sales ranks. More specifically, we run quantile regressions at the 25, 50 and 75% quantile of our dependent variable sales ranks. We estimate the regression model again with fixed product and time effects. Table 5 shows the results for the additional quantile regressions and the effect of the availability of product descriptions.

Table 5. Results of Quantile Regressions for Availability of Product Descriptions

Variable	Estimate 25% quantile	Estimate 50% quantile	Estimate 75% quantile
log(Price)	0.276***	0.289***	0.284***
Average Rating	-0.002	0.007	0.021
log(Number of Reviews + 1)	-0.530***	-0.553***	-0.562***
Amazon Product Description	-0.552***	-0.511**	-0.487**
Manufacturer Product Description	-0.711***	-0.759***	-0.744***

log(Number of Reviews + 1) x Amazon Product Description	0.104*	0.103*	0.101*
log(Number of Reviews + 1) x Manufacturer Product Description	0.057*	0.062**	0.058**
Pencils	-3.235***	-3.180***	-3.128***
Toothbrushes	-2.576***	-2.601***	-2.627***

Significance codes: *** $p < 0.001$, ** < 0.01, * < 0.05

Table 5 indicates that Amazon product descriptions as well as manufacturer descriptions significantly affect sales rank. The influence of manufacturer descriptions is higher than of Amazon descriptions and in opposite to the Amazon descriptions, the influence of manufacturer descriptions slightly increases for niche products. The biggest influence of Amazon product descriptions is for blockbusters. In summary, the effect of the existence of product descriptions is rather robust to products with different sales ranks.

Table 6 depicts the results for the quantile regressions and the effect of the information amount of product descriptions. All findings from Table 4 are supported by the results in Table 6. Thus, they are robust for products with different sales ranks.

Table 6. Results of Quantile Regressions for Information Amount of Product Descriptions

Variable	Estimate 25% quantile	Estimate 50% quantile	Estimate 75% quantile
log(Price)	0.266***	0.277***	0.273***
Average Rating	0.028	0.037*	0.050**
log(Number of Reviews + 1)	-0.629***	-0.645***	-0.660***
log(Length Amazon Description+1)	-0.168***	-0.150***	-0.156***
log(Length Manufacturer Description+1)	-0.144***	-0.153***	-0.153***
log(Number of Reviews + 1) x log(Length Amazon Description+1)	0.045***	0.043***	0.045***
log(Number of Reviews + 1) x log(Length Manufacturer Description+1)	0.012*	0.013*	0.013*
Pencils	-3.189***	-3.136***	-3.084***
Toothbrushes	-2.550***	-2.582***	-2.607***

Significance codes: *** $p < 0.001$, ** < 0.01, * < 0.05

8 Discussion

This paper examined the influence of product descriptions on sales. Based on empirical data from Amazon.com, we found that the existence of product descriptions positively affects sales. This finding is valid for descriptions generated by the manufacturer as well as by Amazon. Amazon descriptions have been found to have a slightly stronger influence on products' sales than descriptions generated by the manufacturer. Products offered online should hence be described by product descriptions.

We furthermore demonstrated that product descriptions that have a higher information amount are more influential on products' sales. The higher the information amount, the better the product description is prepared to reduce consumers' product uncertainty. We demonstrated that especially Amazon product descriptions positively influence sales rank. Finally, we analyzed the interplay between product descriptions and customer reviews and showed that Amazon-generated descriptions especially affect sales of products having no or only a few customer reviews. Manufacturer-generated product descriptions, in contrast, have been found to have a higher impact on sales for products having many reviews. This indicates that consumers might have a higher trust in Amazon descriptions. Further, our robustness check showed that the effect of the existence of product descriptions is independent of its sales rank.

Our research hence has four major managerial implications. First, we analyzed observational data from Amazon. We hence cannot exclude the problem of endogeneity. Manufacturers of successful products (i.e., products with a high number of sales) might have on average longer product descriptions then manufacturers of not successful products. The availability as well as the length of a product description thus can be consequence of the success of a product. Second, manufacturers and retailers should not only incentivize their consumers to provide reviews but also provide product descriptions on their own. We found that an Amazon-generated product description improves a product's sales rank by more than two positions on average.

Third, the longer a product description, the better it helps reducing product uncertainty and the more consumers finally buy the product. Following Mudambi & Schuff [21], we assume that longer product descriptions discuss more product characteristics and hence have a higher information amount.

And fourth, it is worthwhile to provide a product description also for products that already got many customer reviews. In the case of Amazon as online store, we found that especially product descriptions generated by manufacturers have a high impact on a product's sales if there are many customer reviews available.

Our research contributes to the ongoing stream of literature on the impact of information sources on consumers' purchase decisions. We demonstrate that the originator of the product description determines the impact of product descriptions on sales. Amazon-generated descriptions have been found to have a higher impact on sales than descriptions generated by the manufacturer.

Our investigation is subject to three major limitations. First, we analyzed the impact of product descriptions and customer reviews on sales only for three product categories – electric toothbrushes, backpacks and pencils. The influence of product descriptions on sales varies across different product categories. Future research should hence investigate further product categories. Second, we found a significant and positive effect of product descriptions on sales, which indicates that consumers use product descriptions to reduce product uncertainty. We, however, did not measure consumers' product uncertainty and instead assumed that product uncertainty has a strong impact on sales. Investigating the effect of product descriptions on consumers' stated product uncertainty and the effect of the stated product uncertainty on sales provides an interesting avenue for future research. And third, the econometric model we used in this study allows to analyze the correlation between sales and independent variables such as product descriptions and customer reviews. We hence cannot exclude that our results are subject to the problem of endogeneity. It might be the case that good sellers provide a product description whereas bad sellers do not. The difference in sales between sellers providing a product descriptions and sellers that do not is then a result of the sellers' quality rather than the availability of a product description. A seller's quality will be translated into sales. With additional robustness checks we show that our findings are valid for different levels of sales making the endogeneity issue not very likely.

References

1. Akerlof, G.: The Market for " Lemons ": Quality Uncertainty and the Market Mechanism. Q. J. Econ. 84, 488–500 (1970)
2. Dimoka, A., Hong, Y., Pavlou, P.A.: On Product Uncertainty in Online Markets: Theory and Evidence. MIS Q. 36, 395–426 (2012)
3. Overby, E., Jap, S.: Electronic and Physical Market Channels: A Multiyear Investigation in a Market for Products of Uncertain Quality. Manage. Sci. 55, 940–957 (2009)
4. Pavlou, P.A., Liang, H., Xue, Y.: Understanding and mitigating uncertainty in online exchange relationships: a principal- agent perspective. MIS Q. 31, 105–136 (2007)
5. Ghose, A.: Internet exchanges for used goods: an empirical analysis of trade patterns and adverse selection 1. MIS Q. 33, 263–292 (2009)
6. Hong, Y.K., Pavlou, P.A.: Product Fit Uncertainty in Online Markets : Nature , Effects , and Antecedents. Inf. Syst. Res. 25, 328–344 (2014)
7. Ghose, A.., Han, S.P..: Estimating demand for mobile applications in the new economy. Manage. Sci. 60, 1470–1488 (2014)
8. Benlian, A., Titah, R., Hess, T.: Differential Effects of Provider Recommendations and Consumer Reviews in E-Commerce Transactions: An Experimental Study. J. Manag. Inf. Syst. 29, 237–272 (2012)
9. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Inf. Syst. Res. 19, 291–313 (2008)
10. Purnawirawan, N.: Expert Reviewers Beware ! The Effects of Review Set Balance, Review Source and Review Content On Consumer Responses to Online Reviews. J. Electron. Commer. Res. 15, 162–178 (2014)
11. Zhu, F., Zhang, X. (Michael): Impact of Online Consumer Reviews on Sales : The Moderating Role of Product and Consumer. J. Mark. 74, 133–148 (2010)
12. Wang, Z., Huiying, L., Ye, Q., Law, R.: Saliency effects of online reviews embedded in the description on sales: Moderating role of reputation. Decis. Support Syst. 87, 50–58 (2016)
13. Cui, G., Lui, H.-K., Guo, X.: The Effect of Online Consumer Reviews on New Product Sales. Int. J. Electron. Commer. 17, 39–57 (2012)
14. Ehrmann, T., Schmale, H.: The Hitchhiker's Guide to the Long Tail: The Influence of Online-Reviews and Product Recommendations on Book Sales - Evidence from German Online Retailing. Icis. (2008)
15. Banerjee, S., Dellarocas, C., Zervas, G.: Interacting User-Generated Content Technologies: How Questions & Answers Affect Consumer Reviews. Proc. 2017 ACM Conf. Econ. Comput. (2017)

16. Detlor, B., Sproule, S., Gupta, C.: Pre-purchase online information seeking: Search versus browse. J. Electron. Commer. Res. 4, 72–84 (2003)
17. Luo, J., Ba, S., Zhang, H.: The Effectiveness of online shopping characteristics and well-designed websites on satisfaction. MIS Q. 36, 1131–1144 (2012)
18. Liang, T.-P., Huang, J.-S.: An empirical study on consumer acceptance of products in electronic markets: a transaction cost model. Decis. Support Syst. 24, 29–43 (1998)
19. Akdeniz, B., Calantone, R., Voorhees, C.: Effectiveness of Marketing Cues on Consumer Perceptions of Quality: The Moderating Roles of Brand Reputation and Third-Party Information. Psychol. Mark. 30, 76–89 (2013)
20. Dorner, V., Ivanova, O., Scholz, M.: Think Twice Before You Buy! How Recommendations Affect Three-Stage Purchase Decision Processes. Thirty Fourth Int. Conf. Inf. Syst. 5, (2013)
21. Mudambi, S., Schuff, D.: What makes a helpful online review? A Study of Customer Reviews on Amazon.com. MIS Q. 34, 185–200 (2010)
22. Scholz, M., Dorner, V.: The recipe for the perfect review?: An investigation into the determinants of review helpfulness. Bus. Inf. Syst. Eng. 5, 141–151 (2013)
23. Liu, Y.: Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. J. Mark. 70, 74–89 (2006)
24. Archak, N., Ghose, A., Ipeirotis, P.G.: Deriving the Pricing Power of Product Features by Mining Consumer Reviews. Manage. Sci. 57, 1485–1509 (2011)
25. Ghose, A., Ipeirotis, P.G.: Designing Ranking Systems for Consumer Reviews: The Impact of Review Subjectivity on Product Sales and Review Quality. Proc. 16th Annu. Work. Inf. Technol. Syst. (2006)
26. Chevalier, J.A., Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. J. Mark. Res. 43, 345–354 (2006)
27. Clemons, E.K., Gao, G., Hitt, L.M.: When Online Reviews Meet Hyperdifferentiation: A Study of Craft Beer Industry. Proc. 39th Annu. Hawaii Int. Conf. Syst. Sci. 6, 116c–116c (2006)
28. Chen, P.-Y., Wu, S., Yoon, J.: The Impact of Online Recommendation and Consumer Feedback on Sales. Proceeding Int. Conf. Inf. Syst. Paper 58, 711–724 (2004)
29. Kim, Y., Krishnan, R.: On product-level uncertainty and online purchase behavior: An empirical analysis. Manage. Sci. 61, 2449–2467 (2015)
30. Otterbacher, J.: Managing information in online product review communities: A comparison of two approaches. Ecis. (2008)
31. Duan, W., Gu, B., Whinston, A.B.: Do online reviews matter? - An empirical investigation of panel data. Decis. Support Syst. 45, 1007–1016 (2008)
32. Schnapp, M., Allwine, T.: Mining of book data from Amazon.com. Present. UCB/SIMS web Min. Conf. (2001)
33. Chevalier, J., Goolsbee, A.: Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. Quant. Mark. Econ. 1, 203–222 (2003)
34. Ho-Dac, N.N., Carson, S.J., Moore, W.L.: The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? J. Mark. 77, 37–53 (2013)
35. Ghose, A., Ipeirotis, P.G.: Estimating the helpfulness and economic impact of product reviews. Int. J. Innov. Res. Dev. 23, 1498–1512 (2011)
36. Hu, N., Pavlou, P.A., Zhang, J.: On Self-Selection Biases in Online Product Reviews. MIS Q. 41, 449–471 (2017)

Entscheidungsunterstützung durch historienbasierte Dienstreihenfolgeplanung mit Pattern

Lena Wolbeck[1], Bastian Amberg[1], und Natalia Kliewer[1]

[1] Department Wirtschaftsinformatik, Freie Universität Berlin, Berlin, Deutschland
{lena.wolbeck,bastian.amberg,natalia.kliewer}@fu-berlin.de

Abstract. Die Verwendung von Softwarelösungen zur Unterstützung der Personaleinsatzplanung personalintensiver Unternehmen ermöglicht Planungsabläufe zu erleichtern und Einsatzpläne zu optimieren. Dieser Beitrag zielt ab auf eine sozialverträglichere und effizientere Organisation des Personaleinsatzes bei Arbeit im Schichtdienst durch automatisierte Erstellung bestmöglicher Dienstreihenfolgepläne unter Fortschreibung historienbasierter Daten zu Arbeitszeiten und Wunscherfüllung der einzelnen Arbeitskräfte. Als Grundlage dieser Dienstreihenfolgeoptimierung dienen Pattern. Ein Pattern entspricht einem potenziellen Muster an Dienstreihenfolgen in einer Kalenderwoche, das individuell je Arbeitskraft hinsichtlich Gültigkeit, Arbeitszeit und Zufriedenheit bewertet wird. Eingebettet in ein Entscheidungsunterstützungssystem erfolgt die Planerstellung zweistufig. Zuerst werden potenzielle Pattern hinsichtlich der Planungsperiode generiert (Stufe 1). Anschließend erfolgt die Verknüpfung von Pattern und die Generierung eines Dienstreihenfolgeplanes, der den Bedarf an Arbeitskräften abdeckt, mit Hilfe eines Optimierungsmodells (Stufe 2). Der entwickelte Lösungsansatz wird anhand realer Daten eines Wohnheimes für Menschen mit Behinderung evaluiert. Bei maximalen Zufriedenheitswerten generiert das System Dienstreihenfolgepläne, die die Arbeitszeitkonten erheblich besser ausbalancieren als eine manuelle Planung.

Keywords: Entscheidungsunterstützungssystem, Dienstreihenfolge, Pattern, Personaleinsatzplanung, historische Daten

1 Einleitung und Motivation

Jede Einrichtung und jedes Unternehmen, das Dienstleistungen rund um die Uhr (24/7) anbietet, ist mit den Herausforderungen einer lückenlosen Dienstreihenfolgeplanung konfrontiert. Basierend auf längerfristigen Entscheidungen über den Bedarf an Arbeitskräften, die Zusammensetzung der Belegschaft sowie die Art der Schichttypen als Grundlage der Dienste gilt es für den operativen Einsatz der Arbeitskräfte die Dienstreihenfolgeplanung durchzuführen. Eine Dienstreihenfolge bezeichnet hierbei die Abfolge von Diensten und freien Tagen einer Arbeitskraft für eine Planungsperiode.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

1.1 Betrachtete Problemstellung

Bei einem Dienstreihenfolgeplanungsproblem (DRPP) wird folgendes Entscheidungsproblem betrachtet: Es wird eine Zuordnung von einer zu erfüllenden Menge an Diensten einer Planungsperiode zu einer vorgegebenen Menge an Arbeitskräften unter Einhaltung festgesetzter Regeln gesucht. Je nach Anwendungsfall differieren die Zielsetzungen eines DRPP. Unter Einbeziehung individueller Wünsche wird in diesem Beitrag bei der Lösung auf eine maximale Sozialverträglichkeit abgezielt. Quantifiziert wird diese mit Hilfe eines Zufriedenheitswertes sowie der Balance der entstehenden Über- und Unterstunden.

Ohne Beschränkung der Allgemeingültigkeit wird in diesem Beitrag als Planungsperiode ein Monat betrachtet und als Schichttypen zur Dienstbildung werden Frühschicht und Spätschicht berücksichtigt. Jede konkret durchzuführende Schicht entspricht daher einem notwendigen Dienst (pro Tag). Neben Wünschen hinsichtlich bestimmter Dienste bzw. freier Tage werden Urlaubstage und individuelle Arbeitszeitkonten (AZK) der einzelnen Arbeitskräfte beachtet. Bei der Planung sind übliche Regelungen hinsichtlich der Arbeits- und Ruhezeiten wie bspw. *keine Frühschicht nach einer Spätschicht*, *am Wochenende beide Tage frei oder beide Tage Dienst* sowie *Berücksichtigung minimaler und maximaler konsekutiver Arbeitstage* zu beachten.

Die Dienstreihenfolgeplanung ist aufgrund der vielfältigen Regeln eine komplexe Problemstellung und wird heutzutage weitgehend von informationstechnischen Systemen begleitet. Dabei reicht die Entscheidungsunterstützung bei der Planerstellung von simplen Excel-Tabellen zur Veranschaulichung der Pläne zur Erleichterung der manuellen Planung bis hin zu ausgereiften Softwarelösungen mit eingebundenen Optimierungskomponenten, die automatisiert mögliche Pläne vorschlagen. Ein DRPP weist als Optimierungsproblem eine hohe kombinatorische Komplexität auf, die sich aus der hohen Anzahl an möglichen Kombinationen von individuellen Dienstreihenfolgen ergibt. Gemäß [1] ist ein DRPP NP-schwer. Deswegen werden Problemformulierungen häufig vereinfacht, um die Laufzeit der Optimierung zu verringern. Eine simplifizierte Darstellung eines DRPPs kann zu einer Lösung führen, die nicht den realen Anforderungen genügt. Eine historienbasierte Planung, bei der Daten der vorherigen Planungsperiode(n) berücksichtigt werden, wird bspw. oftmals nicht ausgeführt. Die praktische Durchführbarkeit eines Dienstreihenfolgeplanes ist für die Akzeptanz und den Einsatz eines Verfahrens zur Entscheidungsunterstützung jedoch unabdingbar. Eine Anknüpfung an vergangene Perioden erlaubt zudem eine bessere Balance der individuellen AZK und somit eine höhere soziale Verträglichkeit der Pläne.

1.2 Zielsetzung des Beitrages

Dieser Beitrag widmet sich der Fragestellung, wie leitende Arbeitskräfte darin unterstützt werden können, unter Fortschreibung der historischen Daten aus den vorangegangenen Perioden einen optimalen Dienstreihenfolgeplan für eine Planungsperiode zu finden. Der Fokus liegt auf der Erstellung eines Planes, bei dem die Zufriedenheit der Arbeitskräfte sowie ein ausbalanciertes Verhältnis der Über- und Unterstunden innerhalb der Belegschaft als Zielgrößen definiert sind.

Zur Generierung eines optimalen sowie praktikablen Dienstreihenfolgeplanes wird ein zweistufiges Verfahren entwickelt. Die Basis bildet eine Pattern-basierte Planung, die als Instrument für die Planung potenzielle Muster für Dienstreihenfolgen verwendet. Ein Pattern umfasst eine potenzielle Reihenfolge von Diensten und freien Tagen in einem vorab definierten Zeitraum, z. B. eine Woche. Die erste Stufe umfasst die Generierung der potenziellen Pattern. In der zweiten Stufe wird mit Hilfe des entwickelten mathematischen Modells ein optimaler Dienstreihenfolgeplan generiert.

Dieser Ansatz zielt insbesondere auf einen Einsatz in der Praxis ab, weswegen die Berücksichtigung aller praxisrelevanten Restriktionen gewährleistet wird. Als Optimierungskomponente eingebettet in ein IT-System kann dieses Verfahren Dienstreihenfolgepläne empfehlen und die Planung unterstützen. Das mathematische Modell als auch die Implementierung als Entscheidungsunterstützungssystem (engl. *decision support system*, DSS) sind allgemein gehalten, so dass verschiedene zeitliche Planungsperioden sowie andere Schichttypen und Regeln abbildbar sind.

Als Fallbeispiel dient ein Pflegewohnheim für Menschen mit Behinderung. Der entwickelte Ansatz wird anhand dessen realer Daten evaluiert. Die Probleminstanzen für das DRPP umfassen bis zu 53 Arbeitskräfte, die in sechs Wohngruppen organisiert sind. Pro Tag sind bis zu 22 Dienste zu besetzen. Es wird eine heterogene Belegschaft berücksichtigt, da die Arbeitskräfte eine individuell vertraglich vereinbarte Anzahl an Arbeitsstunden pro Woche sowie individuelle Wünsche haben.

Der Beitrag ist wie folgt gegliedert: Abschnitt 2 ordnet diese Arbeit in den Stand der Forschung im Bereich der Dienstreihenfolgeplanung unter Verwendung von Pattern ein. Der entwickelte zweistufige Ansatz wird in Abschnitt 3 (Stufe 1: Pattern-Generierung) und Abschnitt 4 (Stufe 2: Optimierungsmodell) erläutert. Abschnitt 5 skizziert die Einbettung des Lösungsansatzes in ein DSS zur Unterstützung der Planung. Abschnitt 6 diskutiert numerische Ergebnisse eines Fallbeispiels. In Abschnitt 7 wird ein Fazit gezogen und es werden Implikationen für die Praxis abgeleitet.

2 Stand der Forschung – Ansätze für Pattern-basierte Dienstreihenfolgeplanung

Im Kontext der Personaleinsatzplanung gibt es eine Vielzahl an Veröffentlichungen zu Methoden und Lösungsansätzen für die Dienstreihenfolgeplanung, (engl. *rostering/ scheduling*), bei der zunehmend die Berücksichtigung von Wünschen der Arbeitskräfte für die Planbildung in den Fokus rückt. Für einen allgemeinen Literaturüberblick zu „rostering/scheduling“ wird auf [2] und [3] verwiesen.

Bei Problemen der Dienstreihenfolgeplanung sind Restriktionen wie bspw. zur Abdeckung aller benötigten Dienste ähnlich geartet [4]. Differierende Anforderungen und Regelungen hinsichtlich individueller Dienstreihenfolgen erschweren die Verallgemeinerung von Problemmodellierungen und stellen so ein Hindernis bezüglich des Einsatzes von DSS dar, die über eine Optimierungskomponente zur automatisierten Erstellung von Dienstreihenfolgeplänen verfügen. Eine Lösung dieser Schwierigkeit bieten Pattern, da mit ihnen verschiedene Regeln gleich gut abgebildet werden können (wie in Abschnitt 3 verdeutlicht). Da die Mehrzahl an Regeln durch passende

Patternbildung abgedeckt werden können [4], kann ihre Verwendung helfen, ein allgemeingültiges, einsatzfähiges Modell zu formulieren. In diesem Abschnitt liegt der Fokus somit auf exakten und heuristischen Ansätzen, die Pattern als Instrument bei der Dienstreihenfolgeplanung verwenden.

In [5] werden Pattern (engl. auch *stint*) aus konsekutiven Arbeitstagen und freien Tagen jeweils Kostenwerte zugewiesen. Netzwerkbasierte Formulierungen zur Optimierung eines *Nurse Scheduling Problems* mit Zielsetzung der Kostenminimierung liefern für kleine Probleminstanzen gute Ergebnisse. Aufgezeigt wird in [5] zudem, dass eine nicht-zyklische Modellierung von dem Dienstreihenfolgeproblem komplexer ist als ein zyklisches Modell. Basierend auf Pattern ausschließlich für Diensttage werden in [6] zyklische Dienstreihenfolgepläne an mittelgroßen Probleminstanzen mit einer homogenen Belegschaft optimiert. Zielgröße ist hierbei neben Wünschen und Regelungen zur Arbeit am Wochenende die Abfolge von verschiedenen Schichttypen. In [7] werden ebenfalls Pattern bestehend aus Diensten zur Lösung eines kleinen nicht-zyklischen Dienstreihenfolgeproblems genutzt. Diese Pattern sind äquivalent zu möglichen Abfolgen von Schichttypen an aufeinanderfolgenden Tagen. Modelliert als Set Partitioning Problem werden die Anzahl der Dienste bzw. entsprechende Kostenwerte minimiert. Basierend auf denselben Probleminstanzen werden Pattern aus zwei Schichttypen, die einer Kalenderwoche entsprechen, in [8] und [9] genutzt, um jährliche Dienstreihenfolgepläne zu generieren. Die Zielsetzung in diesen beiden Studien ist es, die Erfüllung der Wünsche der bis zu 30 Arbeitskräfte zu maximieren. In einem metaheuristischen Lösungsverfahren von [10] werden monatliche Pattern verwendet, um unter Berücksichtigung von Wünschen für kleine Probleminstanzen automatisiert Dienstreihenfolgepläne zu generieren.

Eine zyklische Planung ist insbesondere bei Planungsproblemen mit kontinuierlichem Bedarf, einer homogenen Belegschaft und Wünschen hinsichtlich Schichtreihenfolgen möglich [6]. Zyklisch zu planen bedeutet zwar die Berücksichtigung vorheriger Planungsperioden, es ist jedoch nicht möglich, Wünsche für spezifische Schichten und einzelne freie Tage zu berücksichtigen. Bei Ansätzen, die nicht historienbasiert sind, d. h. nicht die vorherigen Planungsperioden beachten, können die erzeugten Pläne insbesondere an den Übergängen zwischen den Perioden ungültig sein. Diese Problematik (siehe [7]) führt zu Dienstreihenfolgeplänen, die praktisch nicht einsetzbar sind. Eine fehlende Praktikabilität der mathematischen Lösungsansätze führt zu einer geringen Akzeptanz von Optimierungsmethoden im Bereich der Personaleinsatzplanung [11].

Dieser Beitrag erweitert den aktuellen Forschungsstand, indem der Fokus auf die Integration von historienbasierten Daten bei der Dienstreihenfolgeplanung und damit auf die praktische Einsetzbarkeit einer durch die Verwendung von Pattern verallgemeinerbaren Optimierungskomponente gelegt wird. Diesem identifizierten Problem wird mit einem zweistufigen Verfahren begegnet, das anhand realer Probleminstanzen eines DRPP evaluiert wird. In der ersten Stufe, die in Abschnitt 3 beschrieben wird, werden anhand der Input-Daten potenzielle Pattern generiert. Bei diesem regelbasierten Vorgehen wird eine heterogene Belegschaft berücksichtigt, um die individuell ausführbaren Dienstreihenfolgen sowie die jeweiligen Verknüpfungsmöglichkeiten zwischen den Pattern zu identifizieren. Die ermittelten Mengen an potenziellen Pattern dienen als Grundlage für die im Abschnitt 4 erläuterte zweite Stufe. Da in der ersten Stufe historische Daten explizit betrachtet werden und die zweite Stufe darauf aufbaut, berücksichtigt das weitere Vorgehen Daten der vorherigen Planungsperioden implizit. Mittels mathematischer Optimierung werden in der zweiten Stufe zugleich die individuellen Dienstreihenfolgen ausgewählt und zu einem Plan verknüpft. Das entwickelte DSS erlaubt die Erstellung von historienbasierten, nicht-zyklischen Dienstreihenfolgeplänen unter Berücksichtigung maximaler Sozialverträglichkeit durch Einbeziehen von Wünschen und ausbalancierter Arbeitszeiten.

3 Pattern als Instrument bei der Dienstreihenfolgeplanung

Angelehnt an das Vorgehen einer leitenden Arbeitskraft und aus Gründen der Nachvollziehbarkeit der automatisierten Erstellung eines Dienstreihenfolgeplanes sowie Regelungen hinsichtlich wöchentlicher Arbeitszeiten werden Pattern mit einer Länge von einer Kalenderwoche verwendet. Diese Länge der Pattern wird ebenfalls in [8], [9] und [12] genutzt. Die Zulässigkeit eines Patterns ist individuell je Arbeitskraft bestimmbar. So können beispielsweise Restriktionen hinsichtlich konsekutiver Arbeitstage die individuelle Menge von potenziellen Pattern pro Arbeitskraft verringern. Zudem erfolgt die Bewertung eines Patterns hinsichtlich der Über- und Unterstunden sowie des Zufriedenheitswertes individuell. In diesem Abschnitt wird erläutert, wie potenzielle Pattern individuell bewertet werden (siehe 3.1). In einem ersten Preprocessing-Schritt werden die Pattern basierend auf den Daten einer Probleminstanz hinsichtlich ihrer Zulässigkeit je Arbeitskraft und Woche gefiltert (siehe 3.2). Zudem werden Regeln zu Abhängigkeiten zwischen den Pattern und somit der Verknüpfung der Pattern zwischen den Kalenderwochen im zweiten Preprocessing-Schritt berücksichtigt (siehe 3.3). Dieses Vorgehen ermöglicht in der ersten Stufe des vorgeschlagenen Lösungsansatzes die Generierung von individuellen Patternmengen je Arbeitskraft. Es werden ausschließlich unzulässige Pattern entfernt, so dass keine (heuristische) Reduktion des Lösungsraumes geschieht. Die Komplexität des Entscheidungsproblems wird deutlich reduziert und somit kann die Rechenzeit der Optimierung vermindert werden.

3.1 Dienstreihenfolge als Pattern

Ein Pattern, das einer Dienstreihenfolge von sieben Tagen entspricht, beinhaltet neben Diensten auch freie Tage. In Abbildung 1 ist ein beispielhaftes Pattern dargestellt: Dienste bzw. Schichttypen (Zeile 1) werden den Wochentagen (Zeile 2) zugeordnet. Aufgrund der Betrachtung von zwei Schichttypen, gibt es die folgenden drei Ausprägungen: 'X' ist ein arbeitsfreier Tag, 'F' ist Frühschicht, 'S' ist Spätschicht.

Dienst	X	F	F	S	S	X	X
Wochentag	Mo.	Di.	Mi.	Do.	Fr.	Sa.	So.

Abbildung 1. Beispiel-Pattern

Bewertung der Arbeitszeit: Innerhalb der Pattern werden sowohl Tage, an denen Urlaub gewährt wird, als auch Tage, an denen ein arbeitsfreier Tag zugeteilt wurde, als generell arbeitsfreie Tage mit 'X' gekennzeichnet. Ein freier Tag leistet keinen Beitrag zur Arbeitszeit der Kalenderwoche. Ein Tag mit Urlaub wird bei der Bewertung der Über- bzw. Unterstunden beachtet und in Höhe von bspw. 1/5 der regulären wöchentlichen Arbeitszeit berücksichtigt. Zusätzlich zu Urlaubstagen werden die Arbeitsstunden der Dienste, die in dem Pattern beinhaltet sind, bei der Berechnung der Bewertung beachtet. Die Berechnung der Stundenanzahl eines Pattern sei beispielhaft an dem Pattern aus Abbildung 1 erläutert: Dieses Pattern beinhaltet vier Arbeitstage und drei freie Tage. Wird dieses Pattern anhand der Parameter der Arbeitskraft aus Abbildung 2 berechnet, muss der Urlaubstag 'U' am Montag (Zeile 1) einkalkuliert werden. Angenommen ein Dienst umfasst bspw. acht Stunden und diese Arbeitskraft hat eine wöchentliche Regelarbeitszeit von 38,5 Stunden, würden bei Auswahl dieses Patterns in der entsprechenden Woche 1,2[1] Überstunden anfallen.

Bewertung der Wunscherfüllung bzw. Zufriedenheit: Zusätzlich zu der Bewertung hinsichtlich der Über- bzw. Unterstunden wird einem Pattern ein Zufriedenheitswert abhängig von Arbeitskraft und Woche zugeordnet. Die Bewertung des individuellen Zufriedenheitswertes eines Patterns ist von der Zuteilung zu Diensten bzw. freien Tagen sowie den geäußerten Wünschen der Arbeitskraft abhängig. Liegen für die betrachtete Woche Wünsche für Dienste bzw. nach freien Tagen vor, erhöht ein erfüllter Wunsch den Zufriedenheitswert und ein widersprochener Wunsch hat folgerichtig einen negativen Einfluss. Die Bewertung der Wünsche ist parametrisiert, in der Regel wird einem Tag ein höherer Wert zugewiesen als einer einzelnen Schicht.[2]

Angenommen ein Wunsch eine spezifische Schicht betreffend wird mit (+/-) 10 Einheiten bewertet. Ist hingegen ein gesamter Tag betroffen, wird dieser Wunsch mit (+/-) 20 Einheiten bewertet. Werden keine Wünsche geäußert, wird der Zufriedenheitswert nicht verändert. Als Beispiel für die Bestimmung eines solchen Wertes seien die Wünsche einer Arbeitskraft aus Abbildung 2 verwendet. Für Dienstag ist ein früher Dienst und am Sonntag ein freier Tag gewünscht (Zeile 2). Das Beispiel-Pattern aus Abbildung 1 hat für diese spezifische Woche einen Zufriedenheitswert von +30, da beiden Wünschen entsprochen würde.

Fixierte Dienste/Urlaub	U			S			
Wünsche		F					X
Wochentag	Mo.	Di.	Mi.	Do.	Fr.	Sa.	So.

Abbildung 2. Individuelle fixierte Dienste, Urlaubstage und Wünsche einer Arbeitskraft

3.2 Pattern Preprocessing – Potenzielle Pattern pro Woche und Arbeitskraft

Bei einer Woche von sieben Tagen und drei Ausprägungsmöglichkeiten gibt es insgesamt $3^7 = 2187$ mögliche Pattern ohne Beachtung von Dienstreihenfolge-Regeln.

[1] $4 * 8 + 1/5 * 38{,}5 - 38{,}5 = 1{,}2$

[2] Diese Skalierung kann von einem Anwender des DSS angepasst werden

Da in den Pattern zwischen den Schichttypen und freien Tagen differenziert wird, können Regeln hinsichtlich eines Verbotes der Zuweisung eines Schichttyps an Tag *d* und eines Typs an Tag *d+1* bei der Anzahl der möglichen Pattern berücksichtigt werden. Diese Anzahl kann generell und weiterhin individuell pro Arbeitskraft und Woche bestimmt werden. Dafür werden die folgenden vier Operatoren verwendet:

(1) Kein 'F' nach 'S': Nach einer Spätschicht darf aufgrund der gesetzlich vorgeschriebenen Pausen- und Ruhezeiten keine Frühschicht folgen. In diesem Schritt werden daher alle Pattern, welche die Abfolge 'SF' enthalten, ausgeschlossen. Somit verringert sich die Anzahl der potenziellen Pattern pro Woche auf *567*.

(2a) Wochenende: Um der großen Belastung der Arbeitskräfte aufgrund von Arbeit am Wochenende entgegenzuwirken, gibt es die Regelung, dass entweder beide Tage des Wochenendes (Samstag und Sonntag) Arbeitstage sind oder beide Tage frei sind. Dementsprechend werden die Pattern, die dieser Regelung nicht entsprechen, von den potenziellen Pattern ausgeschlossen. Es sind nunmehr *288* Pattern pro Woche möglich.

(2b) Wochenende-Symmetrie: Zusätzlich zu den unter (2a) ausgeschlossenen Pattern sind Pattern mit der Abfolge 'FS' am Wochenende nicht möglich. Das Ausführen einer Frühschicht am Samstag und einer Spätschicht am Sonntag einer Arbeitskraft würde erzwingen, dass eine andere die konträre Kombination von 'SF' wahrnehmen müsste. Diese Schichtabfolge ist laut (1) nicht möglich. Daher werden weitere 32 Pattern entfernt, es verbleiben *256* potenzielle Pattern pro Woche.

(3) Individuelle konsekutive Arbeitstage: Um die individuellen Mengen an Pattern zu bestimmen, werden persönliche Regelungen zu minimalen und maximalen konsekutiven Arbeitstagen berücksichtigt. Muss eine Arbeitskraft bspw. mindestens zwei konsekutive Arbeitstage leisten, ist die Abfolge 'XFX' sowie 'XSX' verboten und die Anzahl der Pattern reduziert sich auf *161*. Dürfen bspw. nur zwischen zwei und fünf konsekutive Arbeitstage wahrgenommen werden, sind grundsätzlich nur *148* Pattern möglich. Die Beispiele verdeutlichen, dass die Anzahl potenzieller Pattern teilweise drastisch reduziert und daher die Komplexität des Entscheidungsproblems gesenkt werden kann.

Zusätzlich hat die Beachtung der letzten Woche der vorherigen Planungsperiode neben der Sicherstellung der praktischen Einsatzfähigkeit des Dienstreihenfolgeplanes Einfluss auf die Anzahl der individuell möglichen Pattern (siehe Abschnitt 3.3), wodurch die Anzahl an potenziellen Pattern weiter vermindert werden kann.[3] Weiterhin können die potenziellen Patternmengen aufgrund von Urlaubstagen oder anderweitigen Restriktionen zwischen den Wochen differieren.

3.3 Pattern Preprocessing – Abhängigkeiten zwischen den Pattern

Anschließend an die Aufbereitung der individuellen Patternmengen pro Woche werden Abhängigkeiten zwischen den jeweiligen Pattern betrachtet. Hierbei wird ähnlich wie in Schritt (1) und (3) des vorherigen Abschnittes vorgegangen. Die Zuweisung von einem späten Dienst folgend von einem frühen Dienst an zwei konsekutiven Tagen ist

[3] Bei der größten in diesem Beitrag betrachteten Probleminstanz (siehe Abschnitt 6) scheiden so 28,86% aller potenziellen Pattern für die erste Planungswoche aus.

nicht regelkonform. Ein Pattern, das montags mit einem frühen Dienst beginnt, darf nicht nach einem Pattern folgen, das mit einer Spätschicht am Sonntag endet. Darüber hinaus sind minimale und maximale konsekutive Arbeitstage über zwei bzw. drei Pattern hinweg zu beachten. Bei zwei aufeinanderfolgenden Pattern muss die Anzahl der konsekutiven Arbeitstage innerhalb der individuellen Grenzen liegen. Da ein freier Tag eine Folge von konsekutiven Arbeitstagen unterbricht, werden nur die Pattern mit sieben Arbeitstagen (7-Tage-Arbeitswochen, z. B. 'FFFFSSS') für die Abhängigkeiten mit einem vorhergehenden und nachfolgenden Pattern überprüft. Entsprechend der Regeln sind die Pattern und deren Verknüpfungen anzupassen. Abhängig von einem potenziellen Pattern p einer Arbeitskraft e in Woche w werden die nicht erlaubten Pattern für Woche $w+1$ identifiziert (Menge $P_{w+1}^{e,NOT}(p)$). In Kombination mit 7-Tage-Arbeitswochen sind bestimmte Pattern nicht in der vorigen Woche möglich. Für Pattern q (aus der Menge $P_{w-1}^{e}(p)$), die in der Woche $w-1$ vor einer 7-Tage-Arbeitswoche w mit Pattern p möglich sind, sind wiederum gewisse Pattern in der darauffolgenden Woche $w+1$ verboten (Menge $P_{w+1}^{e,NOT}(p,q)$). Als Ergebnis der ersten Stufe dienen diese Mengen als Input für das Optimierungsmodell, das in dem folgenden Abschnitt erläutert wird.

4 Mathematisches Modell zur historienbasierten Dienstreihenfolgeplanung mit Pattern

Als Grundlage für Stufe 2 dient die folgende abstrakte mathematische Modellierung zur Optimierung historienbasierter Dienstreihenfolgeplanung mit der Zielsetzung einer maximalen Sozialverträglichkeit (quantifiziert als Balance der Über- und Unterstunden, sowie dem Gesamtzufriedenheitswert). Ausgangspunkt für die Modellierung bildet die in Abschnitt 1.1 beschriebene Problemstellung. Der Bedarf an erforderlichen Diensten pro Tag ist festgelegt und muss ohne Abweichung erfüllt werden. Zur Verknüpfung mit der Historie liegen individuelle AZK sowie das letzte in der vorherigen Periode durchgeführte Pattern je Arbeitskraft vor. Zusätzlich ist die Anzahl der konsekutiven Arbeitstage am Ende der letzten Periode bekannt, sollte diese sieben überschreiten.

Für die aktuelle Planungsperiode sind neben dem Bedarf Wünsche, Urlaubstage und fixierte Dienstzuteilungen der Arbeitskräfte bekannt. Diese werden wie die zahlreichen Regelungen zu Dienstreihenfolgen mit Hilfe der individuellen Mengen an Pattern pro Kalenderwoche berücksichtigt. Wesentlich für die Modellbildung ist die automatisierte Generierung der individuellen Patternmengen als Eingabe für das Optimierungsproblem (Stufe 1). Durch die Festlegung dieser Mengen ergeben sich verschiedene Kombinationsmöglichkeiten von Dienstreihenfolgen und somit verschiedene zulässige Planlösungen. Einerseits existieren die generell pro Arbeitskraft und Woche möglichen Pattern (ermittelt gemäß Abschnitt 3.2), mit „7-Tage-Arbeitswoche“-Pattern als besondere Teilmenge. Zusätzlich existieren Patternmengen (ermittelt gemäß Abschnitt 3.3), die in Abhängigkeit der einzelnen Pattern einer Woche und Arbeitskraft mögliche oder verbotene wochenübergreifende Verknüpfungen zu anderen Pattern regeln.

Die Formulierung des mathematischen Modells erfolgt als gemischt-ganzzahliges Optimierungsproblem (engl. *Mixed-Integer Programming*, MIP) mit binären Entscheidungsvariablen über die konkrete Auswahl jedes möglichen Pattern und reellen Entscheidungsvariablen zur Abbildung der Zielfunktionsgrößen. Tabelle 1 fasst Mengen, Parameter und Entscheidungsvariablen des Modells zusammen.

Tabelle 1. Mengen, Parameter und Entscheidungsvariablen des Modells

Mengen:	
E	Menge der Arbeitskräfte, mit Index e
W	Menge der Wochen, mit Index w, {0, ... |W|}
D_w	Menge von Tagen, mit Index d, innerhalb Woche w
S	Menge der Schichttypen, mit Index s
P_w^e	Menge von Pattern für Arbeitskraft e in Woche w, Menge der Woche 0 besteht nur aus dem letzten durchgeführten Pattern
$\tilde{P}_w^e$	Menge der „7-Tage-Arbeitswoche" Pattern für Arbeitskraft e in Woche w, Teilmenge von P_w^e
$P_{w+1}^{e,NOT}(p)$	Menge von Pattern für Arbeitskraft e, die nicht in Woche $w+1$ zugewiesen werden können, aufgrund der Auswahl von Pattern p in Woche w
$P_{w-1}^e(p)$	Menge von Pattern für Arbeitskraft e, die in Woche w-1 zugewiesen werden können, in Kombination mit „7-Tage-Arbeitswoche" Pattern p aus Woche w
$P_{w+1}^{e,NOT}(p,q)$	Menge von Pattern für Arbeitskraft e, die nicht in Woche $w+1$ zugewiesen werden können, aufgrund der Auswahl von „7-Tage-Arbeitswoche" Pattern p in Woche w und der Auswahl von Pattern q in Woche w-1
Parameter:	
ϵ_{sat}	Mindestwert an Zufriedenheit, der erreicht werden soll
α	Bestrafungsfaktor in Zielfunktion für Abweichung von ϵ_{sat}
dem_{ds}	Bedarf an Diensten vom Schichttyp s an Tag d
sat_{ep}	Zufriedenheitswert von Pattern p für Arbeitskraft e, umfasst für das Pattern in der Woche 0, die bisher kumulierten Werte
eh_{ep}	Summe der Über- oder Unterstunden durch Pattern p für Arbeitskraft e, umfasst für das Pattern in der Woche 0, die bisher kumulierten Werte
a_{epds}	1, falls Pattern p von Arbeitskraft e den Schichttyp s an Tag d abdeckt, sonst 0
Entscheidungsvariablen:	
x_{ep}	1, falls Arbeitskraft e Pattern p ausüben soll, sonst 0 (binär)
u	Diskrepanz um vorgegebenen Zufriedenheitswert zu erfüllen (reelle Zahl)
eh_{max}	Höchste Überstundenanzahl von allen Arbeitskräften (reelle Zahl)
eh_{min}	Höchste Unterstundenanzahl von allen Arbeitskräften (reelle Zahl)

Zur gleichzeitigen Verfolgung beider Zielsetzungen zur Sozialverträglichkeit ist die Formulierung eines bikriteriellen Optimierungsproblems notwendig. Bei solchen Problemstellungen existiert in der Regel keine eindeutig optimale Lösung, sondern es gibt eine Menge von sogenannten „effizienten Lösungen". Diese können nicht verbessert werden, ohne dass sich die Ausprägung einer Zielsetzung verschlechtert (vgl. [13]). Die Menge effizienter Lösungen bildet somit die möglichen Kompromisse zwischen der Erfüllung der beiden Zielsetzungen ab. Sie kann mittels der ϵ-Constraint-Methode bestimmt werden, bei der die Ziele gemäß einer Rangfolge geordnet und aufeinander aufbauend einzelne Optimierungsprobleme gelöst werden. Die zuvor erreichte Zielsetzung wird dabei zu einer Nebenbedingung umformuliert, die im neuen Optimierungsproblem zu einem Mindestanteil, vorgegeben durch den ϵ-Wert, erfüllt werden muss. In dieser Arbeit wird zunächst ein Optimum für die Zielsetzung nach maximaler Zufriedenheit berechnet. Anschließend wird als zweite Zielsetzung die Ausgeglichenheit der Über- und Unterstunden unter der Voraussetzung optimiert, dass die maximal mögliche Zufriedenheit zu einem Mindestwert ϵ_{sat} erfüllt wird. Damit das entstehende MIP rechnerisch einfacher lösbar ist, wird dies als Nebenbedingung durch eine Elastic-ϵ-Constraint formuliert (vgl. [13]), bei der eine zusätzliche Diskrepanz-Variable mögliche Abweichungen auffangen kann. Durch Veränderung des zu erfüllenden Mindestwertes sind systematisch alle effizienten Lösungen bestimmbar. Beide Zielsetzungen sind gleichzeitig vollständig erfüllbar, wenn der maximal mögliche Zufriedenheitswert zu 100% (ohne Einsatz der Diskrepanz-Variable) erreicht wird und zugleich der berechnete Zielfunktionswert des Gesamtproblems den gleichen Wert wie bei Zielsetzung nach reiner Optimierung der Ausgeglichenheit annimmt.

Das mathematische Optimierungsmodell zur Dienstreihenfolgeplanung unter Fortschreibung historischer Daten vergangener Planungsperioden (*h-DRPP*$^{(eh,sat)}$) und zweifacher Zielsetzung kann wie folgt formuliert werden:

$$Minimiere\ (eh_{max} - eh_{min}) + \alpha * u \tag{1}$$

$$s.t \quad \sum_{e \in E} \sum_{w \in W} \sum_{p \in P_w^e} sat_{ep} x_{ep} + u \geq \epsilon_{sat} \tag{2}$$

$$\sum_{e \in E} \sum_{p \in P_w^e} a_{epds} x_{ep} = dem_{ds} \quad \forall w \in W \setminus \{0\}, \forall \mathrm{d} \in D_w, \forall s \in S \tag{3}$$

$$\sum_{p \in P_w^e} x_{ep} = 1 \qquad \forall e \in E, \forall w \in W \tag{4}$$

$$x_{ep} + \sum_{r \in P_{w+1}^{e,NOT}(p)} x_{er} \leq 1 \qquad \forall e \in E, \forall w \in W \setminus \{|W|\}, \forall p \in P_w^e \tag{5}$$

$$x_{ep} + x_{eq} + \sum_{r \in P_{w+1}^{e,NOT}(p,q)} x_{er} \leq 2 \ \ \forall e \in E, \forall w \in W \setminus \{0\}, \forall p \in \tilde{P}_w^e, \forall q \in P_{w-1}^e(p) \tag{6}$$

$$eh_{max} \geq \sum_{w \in W} \sum_{p \in P_w^e} eh_{ep} x_{ep} \quad \forall e \in E \tag{7}$$

$$eh_{min} \leq \sum_{w \in W} \sum_{p \in P_w^e} eh_{ep} x_{ep} \quad \forall e \in E \tag{8}$$

$$x_{ep} \in \{0,1\} \quad \forall e \in E, \forall w \in W, \forall p \in P_w^e \tag{9}$$

$$u, eh_{max}, eh_{min} \in \mathbb{R} \tag{10}$$

Mit der Zielfunktion (1) wird die Differenz zwischen der höchsten Überstundenzahl und der höchsten Unterstundenzahl von allen Arbeitskräften minimiert und gleichzeitig werden mögliche Abweichungen vom zu erfüllenden Zufriedenheitswert minimiert. Die Restriktion (2) stellt dabei sicher, dass ein Mindestmaß an vorgegebener Gesamtzufriedenheit erfüllt wird. Die Schwere möglicher Abweichungen von dieser Vorgabe wird in Abhängigkeit des Bestrafungsfaktors α in der Zielfunktion gesteuert.

Die Restriktionen (3) stellen die Bedarfserfüllung pro Tag und Schichttyp sicher. Restriktionen (4) gewährleisten, dass pro Arbeitskraft genau ein Pattern pro Woche ausgewählt wird. Mit den Restriktionen (5) wird die Durchführung von Pattern in aufeinanderfolgenden Wochen vermieden, die nicht miteinander verknüpft werden dürfen. Restriktionen (6) stellen sicher, dass die maximale Anzahl aufeinanderfolgender Arbeitstage pro Arbeitskraft eingehalten wird. Restriktionen (7) bzw. (8) bestimmen die maximale bzw. minimale Anzahl zu leistender Zusatzstunden der Arbeitskräfte. Restriktionen (9) und (10) legen die Wertebereiche der Entscheidungsvariablen fest.

Vorgaben zu dem Zufriedenheitswert ϵ_{sat} können durch ein abgewandeltes, vereinfachtes Modell berechnet werden. Dazu wird (1) durch die Zielfunktion (11) ersetzt. In Kombination mit den Restriktionen (3), (4), (5), (6) und (9) entsteht ein Optimierungsmodell (*h-DRPP*$^{(sat)}$), das ausschließlich die Gesamtzufriedenheit aller Arbeitskräfte maximiert. Der resultierende Dienstreihenfolgeplan ist für den Einsatz zulässig, ohne jedoch auf einen Ausgleich der Über- und Unterstunden abzuzielen. Die Lösung kann als Vorgabe für obiges Modell *h-DRPP*$^{(eh,sat)}$ einen maximal möglichen Zufriedenheitswert liefern, der entsprechend der Wahl von ϵ_{sat} berücksichtigt wird.

$$Maximiere\ \sum_{e\in E}\sum_{w\in W}\sum_{p\in P_w^e} sat_{ep}x_{ep} \qquad (11)$$

Weitere Modellvariationen: Für eine Dienstreihenfolgeplanung mit alleiniger Zielsetzung nach Ausgeglichenheit von Über- und Unterstunden (*h-DRPP*$^{(eh)}$) entfällt in der Formulierung (1)-(10) die Restriktion (2) und der korrespondierende Summand $\alpha * u$ in der Zielfunktion (1). Für eine Dienstreihenfolgeplanung ohne Verknüpfung zu vergangenen Planungsperioden kann die Formulierung (1)-(10) wie folgt zu einem Modell *DRPP*$^{(eh,sat)}$ angepasst werden: Als Eingabeparameter verändert sich die Wochenmenge W in die Menge $\{1,\ldots,|W|\}$. Zusätzlich sind die Restriktionen (6) über alle Wochen aus $W\backslash\{1\}$ statt aus $W\backslash\{0\}$ zu formulieren.

5 Einbettung in ein Entscheidungsunterstützungssystem

Der entwickelte Lösungsansatz kann wie in Abbildung 3 dargestellt in ein DSS eingebettet und für die Dienstreihenfolgeplanung verwendet werden (orientiert an der aktuellen Planung des in diesem Beitrag betrachteten Fallbeispiels). Als Eingabe aggregiert die leitende Arbeitskraft die notwendigen Daten zu Bedarf an Diensten, Arbeitskräften und deren Restriktionen sowie weiterer Regelungen in einer Excel-Datei. Die manuelle Planung wird anschließend durch das entwickelte Verfahren inklusive Optimierung ersetzt. Diese Optimierungskomponente ist in C# implementiert und verwendet als Solver CPLEX. Auf Grundlage der potenziellen Patternmengen werden optimale Dienstreihenfolgepläne generiert und ausgegeben – inklusive eines Lösungsreportes über den Optimierungsprozess. Der Datenaustausch zwischen den einzelnen Komponenten erfolgt durch strukturierte Textdateien (.csv). Die Lösungsgüte der Pläne kann neben einer tabellarischen Darstellung auch in einem Koordinatensystem im Hinblick auf die beiden Zielsetzungen visualisiert und miteinander verglichen werden. Schlussendlich kann aus den vorgeschlagenen Plänen ein geeigneter Plan ausgewählt und bei Bedarf manuell angepasst werden.

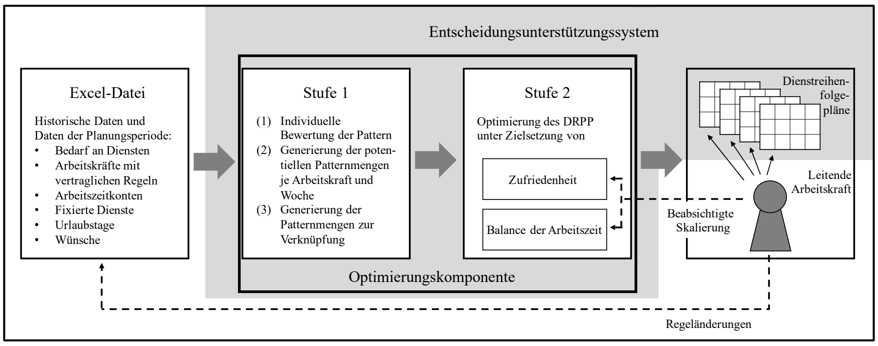


Abbildung 3. Entscheidungsunterstützungssystem für die Planung

Optimierungsbasierte DSS erfordern in der Regel eine aufwendige Einführungsphase. Im Rahmen von Einführungsprojekten wird auch eine Kalibrierung der Optimierungskomponenten vorgenommen. Gegenstand ist bspw. die Bewertung der Wünsche. Zudem wird dem Anwender des DSS als Interaktion ermöglicht, eine Skalierung zwischen Plänen mit maximalem Zufriedenheitswert und bestmöglicher Balance der Über- und Unterstunden auszuwählen (inklusive der Gewichtung möglicher Abweichungen von der gewünschten Zufriedenheit durch α) sowie ein Zeitlimit für die Optimierung vorzugeben. Anwendungsspezifisch eingestellt kann das entwickelte DSS dementsprechend in verschiedenen Fällen eingesetzt werden.

6 Numerische Ergebnisse eines Fallbeispiels

Der vorgestellte zweistufige Lösungsansatz zur Dienstreihenfolgeplanung wird anhand realer Daten eines Pflegewohnheims für Menschen mit Behinderung evaluiert. Dabei liegt der Fokus neben dem Vergleich historienbasierter Planung mit einer Planung ohne Bezug zur Vorperiode auf der Ermittlung des gesamten Potenzials optimierungsbasierter Planung. Das Pflegewohnheim umfasst sechs Wohngruppen, für die bisher einzelne Dienstreihenfolgepläne erstellt wurden. Eine Verteilung der Arbeitskräfte zwischen den Gruppen ist möglich bzw. häufig notwendig, wurde aber bisher aufgrund der nicht zu bewältigenden Komplexität einer manuellen Planung nicht ausgeführt. Generell sind im Wohnheim pro Tag 22 Dienste zu besetzen. Dabei ist eine heterogene Belegschaft zu berücksichtigen, da die insgesamt 53 Arbeitskräfte individuell vertraglich vereinbarte Wochenarbeitszeiten haben sowie individuelle Wünsche für Arbeits- bzw. Urlaubszeiten für die Planungsperiode geäußert werden können. Die Planungsperiode umfasst einen Monat. In diesem Beitrag werden 14 Probleminstanzen zur Dienstreihenfolgeplanung eines Monats betrachtet. Zum einen werden Dienstreihenfolgepläne separat für die sechs Wohngruppen erstellt wie es der bisherigen manuellen Planung entspricht. Einer Wohngruppe sind jeweils sechs bis elf Arbeitskräfte zugeordnet, pro Gruppe und Tag sind (mit einer Ausnahme) generell vier Dienste zu planen. Zum anderen wird ein gemeinsamer Dienstreihenfolgeplan über alle sechs Gruppen erstellt, um mögliches Potenzial aufzudecken, das bisher nicht erschließbar ist. Für diese sieben Instanzen werden Pläne sowohl historienbasiert als auch ohne Bezug zur Vorperiode erstellt (insgesamt 14). Für jede der Probleminstanzen werden mit den Modellen *(h-)DRPP*$^{(sat)}$ und *(h-)DRPP*$^{(eh)}$ die jeweils bestmögliche Ausprägung der Zielsetzung nach maximaler Wunscherfüllung und nach minimaler Abweichung von Über- und Unterstunden berechnet. Mit dem Modell *(h-)DRPP*$^{(eh,sat)}$ werden dann beide Zielsetzungen gleichzeitig verfolgt, wobei als generelle Vorgabe der maximale Zufriedenheitswert erreicht werden soll (mögliche Verletzungen werden mit α=1 bei der Berechnung des Zielfunktionswertes (1) bestraft, die Zusammensetzung der Zufriedenheitswerte pro Pattern ist skaliert wie in Abschnitt 3.1 beschrieben).[4]

Tabelle 2 zeigt die Zielfunktionswerte hinsichtlich Zufriedenheit sowie Ausgeglichenheit der Über- und Unterstunden exemplarisch für eine einzelne Wohngruppe (*planOpt_06*) sowie für das gesamte Wohnheim (*planOpt_gesamt*) bei historienbasierter Planung (*mit h*) bzw. bei Planung ohne Bezug zur Vorperiode (*ohne h*).

Tabelle 2. Zufriedenheitswerte sowie minimierte Differenz von Über-/Unterstunden in hh:mm

Instanz	*(h-)DRPP*$^{(sat)}$	*(h-)DRPP*$^{(eh)}$		*(h-)DRPP*$^{(eh,sat)}$		
	mit h und ohne h	*mit h*	*ohne h*	*mit h*	*ohne h*	*bei* ϵ_{sat}
planOpt_06	**490**	**112:17**	17:42	120:36	17:42	(100%) **490**
				112:17	17:42	(90%) 450
planOpt_gesamt	**1340**	**277:58**	30:52	**277:58**	30:52	(100%) **1340**

In den Tests hat sich gezeigt, dass (bis auf eine Ausnahme) für alle Instanzen gleichzeitig der höchste Zufriedenheitswert[5] bei minimal möglicher Differenz von eh_{max} und eh_{min} erreichbar ist. Wie in Tabelle 2 zu erkennen, gilt diese Beobachtung auch für die alle Wohngruppen umfassende Gesamtplanung (*planOpt_gesamt*). Ausnahme bildet die ebenfalls in Tabelle 2 aufgeführte historienbasierte Einzelplanung für die Wohngruppe 6, bei der neben der Historie vergleichsweise viele Wünsche der Arbeitskräfte im Planungszeitraum zu berücksichtigen sind. Der maximale Zufriedenheitswert von 490 (100%) ist nur mit einer leichten Erhöhung des minimierten Differenzwertes der Über- und Unterstunden erreichbar. Anders ausgedrückt, der optimale Zielfunktionswert der Über- und Unterstunden ist erst bei 90% Zufriedenheit erreichbar.

Anzumerken ist, dass die ohne Historie erstellten Pläne zwar sehr ausgeglichen mit in der Regel geringen Unterschieden in den individuellen AZK der Arbeitskräfte sind, bei einer Verknüpfung mit der Vorperiode allerdings zu Regelverletzungen führen

[4] Die Testläufe wurden auf einem Notebook mit Intel Core i7 Prozessor mit 2,6 Ghz und 16 GB Hauptspeicher durchgeführt. Für die historienbasierte Planung liegen alle Laufzeiten in der Regel unter 3 Minuten. Für reine Zufriedenheitsmaximierung liegen die Zeiten im geringen Sekundenbereich. Bei Planung ohne Bezug zur Historie kann der Optimalitätsbeweis beim Lösen andauern, die Laufzeiten liegen dabei in der Regel deutlich unter 30 Minuten. In allen Fällen wurden vom Solver bereits zu Beginn des Lösungsprozesses zulässige Pläne gefunden. Ein vom Anwender ausgeführter frühzeitiger Abbruch der Optimierung durch striktes Zeitlimit kann daher zwar nicht unbedingt zu optimalen, aber zu passablen Lösungen führen.

[5] Da keine Daten über die bisherige Wunscherfüllung vorliegen, sind die maximalen Zufriedenheitswerte bei Planung *mit h* und *ohne h* gleich.

würden. Bei separater Planung ergeben sich im Hinblick auf verbotene 'SF'-Wechsel, sowie einzuhaltende maximale und minimale konsekutiver Arbeitstage insgesamt 8 Regelverletzungen, bei Wohngruppen übergreifender Gesamtplanung sind es 10.

Tabelle 3. Ausgeglichenheit von Über-/Unterstunden bei manuellen und optimierten Plänen

Über- u. Unterstunden (Eh)	*vorher*	*planMan*	*planOpt_einzeln*	*planOpt_gesamt*
Maximum (eh_{max})	139:57	(+4%) 145:17	(-57%) 59:39	(-82%) 25:32
Minimum (eh_{min})	-356:02	(+34%) -478:18	(-29%) -252:26	(-29%) -252:26
Differenz von Max - Min	495:59	(+26%) 623:35	(-37%) 312:05	(-44%) 277:58
Summe Überstunden +	1575:03	(+16%) 1829:44	(-49%) 809:12	(-69%) 479:57
Summe Unterstunden -	-860:40	(+32%)-1140:33	(+72%)-1480:56	(+25%)-1079:41
Gesamtstunden +/-	2435:43	(+22%) 2970:17	(-6%) 2290:08	(-36%) 1559:38
Mittelwert je Arbeitskraft	13:29	13:00	-12:40	-11:19
Standardabweichung	68:35	87:50	55:40	48:32

Tabelle 3 vergleicht die Ergebnisse von Varianten historienbasierter Dienstreihenfolgeoptimierung sowie der manuellen Planung. Da bei der manuellen Planung sämtliche Wünsche erfüllt werden und bei Optimierung der maximale Zufriedenheitswert erreicht werden kann, liegt der Fokus auf Kennzahlen zur Ausgeglichenheit der AZK (*Über- u. Unterstunden*). Ausgehend von kumulierten Daten der Vorperioden (*vorher*) werden diese für manuelle Planung (*planMan*) sowie für die – zusammengefasste – separate Optimierung aller Wohngruppen (*planOpt_einzeln*) und die Wohngruppen übergreifende Gesamtoptimierung (*planOpt_gesamt*) aufgeschlüsselt mit jeweiligen prozentualen Veränderungen in Bezug auf die Vorperiode. Es wird deutlich, dass durch die Optimierung eine signifikante Steigerung der Sozialverträglichkeit erreicht werden kann. Vor dem Hintergrund maximaler Zufriedenheitswerte werden Überstunden drastisch abgebaut. Zusätzlich zeigt sich, dass die bisher nicht umsetzbare Gesamtplanung erhebliche Potenziale verglichen mit separater Planung besitzt.

7 Fazit und Ausblick

Dieser Beitrag beschreibt aktuelle Entwicklungen für die informationstechnische Unterstützung der Personaleinsatzplanung mit einem Fokus auf Praktikabilität der Lösungsverfahren und Akzeptanz der automatisiert erstellten Lösungen. Es wurde ein zweistufiges Verfahren zur Lösung eines DRPP vorgestellt, das insbesondere auf die Sozialverträglichkeit der erstellten Pläne abzielt und eine historienbasierte Anknüpfung an vorausgegangene Perioden ermöglicht. Nach der problemspezifischen Generierung der Mengen potenzieller Pattern können mit Hilfe eines Optimierungsmodells je nach Skalierung der Zielsetzungen verschiedene bestmögliche Pläne erstellt werden. Darüber hinaus ist das entwickelte Optimierungsmodell für verschiedene Varianten eines DRPP einsetzbar und leicht anpassbar, bspw. für andere Planungszeiträume, Schichttypen und Regeln sowie unterschiedliche Längen der Pattern.

Die Anwendbarkeit und der Nutzen des Ansatzes wurden evaluiert und dargestellt. Der Lösungsansatz liefert für ein Fallbeispiel sehr gute Ergebnisse. Unter Beibehaltung hoher Zufriedenheitswerte ist parallel eine deutliche Reduzierung der Überstunden möglich. Insgesamt zeigt sich, dass eine bessere Balance der Über- und Unterstunden bei einer historienbasierten Planung bewirkt werden kann. Es können darüber hinaus organisatorisch komplexere Problemstellungen betrachtet und zusätzliche Freiheitsgrade bei der Planung genutzt werden. Eine empirische Langzeitstudie kann die Auswirkungen der durch Optimierung unterstützten Planung auf die Belegschaft untersuchen. Da das DSS Wünsche leichter berücksichtigen kann als die manuelle Planung, wird erwartet, dass die Wunschanzahl im Zeitverlauf steigt.

Der Einfluss und das Zusammenspiel von Skalierung und Diversität der Wünsche sowie der Verschiedenartigkeit vorhandener AZK auf eine gute oder schlechte Lösbarkeit historienbasierter DRPP ist darüber hinaus als grundlegende Fragestellung weiter zu untersuchen. Hierfür und für einen möglichen Vergleich mit problem-spezifischen Heuristiken und metaheuristischen Ansätzen bietet das vorgestellte Modell und seine Anwendung im Rahmen eines DSS einen passenden Ansatzpunkt.

Literaturverzeichnis

1. Osogami, T., Imai, H.: Classification of various neighborhood operations for the nurse scheduling problem. In: International Symposium on Algorithms and Computation. 2000. Springer.
2. Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., De Boeck, L.: Personnel scheduling: A literature review. European Journal of Operational Research, 2013. 226(3): S. 367-385.
3. Ernst, A.T., Jiang, H., Krishnamoorthy, M., Sier, D: Staff scheduling and rostering: A review of applications, methods and models. European Journal of Operational Research, 2004. 153(1): S. 3-27.
4. Burke, E.K., Curtois, T.K New approaches to nurse rostering benchmark instances. European Journal of Operational Research, 2014. 237(1): S. 71-81.
5. Millar, H.H., Kiragu, M.: Cyclic and non-cyclic scheduling of 12 h shift nurses by network programming. European Journal of Operational Research, 1998. 104(3): S. 582-592.
6. Becker, T., Steenweg, P.M., Werners, B.: Cyclic shift scheduling with on-call duties for emergency medical services. Health Care Management Science, 2018: S. 1-15.
7. Valouxis, C., Housos, E.: Hybrid optimization techniques for the workshift and rest assignment of nursing personnel. Artificial Intelligence in Medicine,2000. 20(2): p.155-175.
8. Aickelin, U., Dowsland, K.A.: Exploiting problem structure in a genetic algorithm approach to a nurse rostering problem. Journal of Scheduling, 2000. 3(3): S. 139-153.
9. Aickelin, U., Dowsland, K.A.: An indirect genetic algorithm for a nurse-scheduling problem. Computers & Operations Research, 2004. 31(5): S. 761-778.
10. Ikegami, A., Niwa, A.: A subproblem-centric model and approach to the nurse scheduling problem. Mathematical Programming, 2003. 97(3): S. 517-541.
11. Glass, C.A., Knight, R.A.: The nurse rostering problem: A critical appraisal of the problem structure. European Journal of Operational Research, 2010. 202(2): S. 379-389.
12. Hadwan, M., Ayob, M.: A constructive shift patterns approach with simulated annealing for nurse rostering problem. In: 2010 International Symposium in Information Technology (ITSim), 2010. IEEE.
13. Ehrgott, M., Wiecek, M.M.: Mutiobjective programming. Multiple Criteria Decision Analysis: State of the Art Surveys, 2005. Springer. 78: S. 667-708.

A Semi-Automated Approach for Generating Online Review Templates

Martin Poniatowski[1], Jürgen Neumann[1], Thomas Görzen[1], and Dennis Kundisch[1]

[1] Paderborn University, Business Administration and Economics, Paderborn, Germany
{martin.poniatowski,juergen.neumann,thomas.goerzen,
dennis.kundisch}@wiwi.uni-paderborn.de

Abstract. It is widely accepted that online customer reviews play a key role in influencing potential customers. Online shops and rating platforms face two important challenges: how to elicit reviews from existing customers, and how to maximize the helpfulness of these reviews. To address these challenges, online review systems require a design that facilitates the writing process and increases the reviews' informational value. It is an open question whether providing reviewers with review templates could support both of these objectives. In this study we propose a semi-automated approach for designing review templates, which uses machine-learning for the generation and crowdworkers for the evaluation of review templates. To identify product- or service-specific dimensions we use Latent Dirichlet Allocation on reviews considered helpful by readers. Further research is planned to establish whether the automated approach for deriving and evaluating templates is an effective method for improving review generation for specific products and services.

Keywords: Online Reviews, Review Template, Review Writing Facilitation, Topic Modelling, Latent Dirichlet Allocation

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Machine Learning goes Measure Management: Leveraging Anomaly Detection and Parts Search to Improve Product-Cost Optimization

Matthias Walter[1], Volodymyr Vasyutynskyy[2], Duc Anh Trinh[1], and Christian Leyh[1]

[1] Technische Universität Dresden, Chair of Information Systems, esp. IS in Manufacturing and Commerce, Dresden, Germany
{matthias.walter3,duc_anh.trinh,christian.leyh}@tu-dresden.de
[2] SAP SE, Innovation Center Network, Dresden, Germany
{volodymyr.vasyutynskyy}@sap.com

Abstract. In many industries, particularly discrete manufacturing, companies can benefit from conducting product-cost optimization as early as possible. Given the amount of data to be analyzed in the costing process, the lack of dedicated information system support, and the pressure to quickly estimate the cost of new products, the potential for cost optimization is often underexploited. In this paper, we present an approach for leveraging machine learning capabilities, including similarity and anomaly analysis, to improve the identification of product-cost optimization potentials and therefore, improve the quality of early cost estimates. For the approach to succeed, however, ongoing training of a model based on a high-quality dataset is crucial. Thus, we also propose the machine learning approach's integration with our long-term research project toward improving the management of cost optimization during product development.

Keywords: Product costing, product-cost optimization, machine learning, product development, measure management.

1 Problem Context

To ensure long-lasting economic success amid globalization, shortened product life cycles, and growing product diversity, companies seek support for conducting product-cost optimization [1], [2]. Despite the immense potential of early cost optimization [3], [4], our prior research has shown that early optimization lacks dedicated information system (IS) support [2], [5]. In response, our long-term research project grounded on design science research methodology [6], [7] aims to improve IS support for early cost optimization. Based on industry-evaluated requirements and IS implementation hurdles, we developed an approach that aims at improving the management of cost-optimization measures, supporting the phases of identifying, elaborating, evaluating, and implementing optimization measures. This measure management (MM) approach, including optimization examples, is depicted in [5] and [8] in detail.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

In this context, this paper takes up an idea that emerged from our research project: The automated identification of optimization potentials [2], [8]. Such a functionality is especially important in the discrete manufacturing industry, where products such as cars and special-purpose machines comprise up to 100,000 components. Due to this and further practical challenges [2], [8], the complexity of optimization is hardly manageable manually. Therefore, industry demands IS approaches to support the identification of optimization potentials by using available datasets [5], [9].

Thus, on the one hand, this paper showcases machine learning (ML) approaches for identifying optimization potentials. On the other hand, the paper identifies a means of improving the accuracy of ML algorithms by incorporating the MM approach from our long-term research project.

2 Machine Learning goes Measure Management

2.1 Measure Identification

To demonstrate the potential of ML for early product-cost optimization, we first introduce two exemplary optimization use cases from industry [5]. We use these to develop suitable approaches for self-optimizing algorithms. These approaches then form the foundation for demonstrating the interaction between the MM approach based on [8] and self-optimizing algorithms that support the identification of cost optimization potentials. This paper summarizes such self-optimizing algorithms under the term ML according to the definition in [10], which defines ML as a specific system design that allows the system to learn from data and improve with experience [10].

The first approach, a similar parts search, helps users locate similar but preferable (e.g., cheaper) materials [9]. For that, a user selects a certain material in a product costing structure. Based on dedicated material attributes like description, material number, weight, plant, etc., the approach starts the search for materials with similar attributes in the master data and previous costing structures. The single similar attributes are found using fuzzy search [11], with subsequent defining of the overall degree of similarity of the target and similar materials as a weighted sum of single similarity scores. The weight coefficients are defined using an optimization algorithm which analyzes the historical similar products. As result, the ML algorithm recommends a list of similar materials, ranked by their degree of similarity to the one selected [12]. This list of similar but possibly more suitable materials can be used to create an optimization measure as described in [5]. By creating such a measure, the user can accept the respective recommendation to replace the originally selected material.

The second approach involves plausibility checks, based on the work of Vosough and Vasyutynskyy [13], whose main goal was identifying potential user-made errors by assessing the validity and plausibility of items within a costing structure. Since only a small set of examples of such errors is known, which makes usage of the supervised ML approaches difficult, Vosough and Vasyutynskyy concentrated on the unsupervised model-based algorithms for anomaly detection. These algorithms detect potential errors in cost estimates as anomalies, i.e., significant deviations from the values of typical material attributes (e.g., price and weight) or typical costing structures, in which the attribute values and structures are learned from historical cost estimates. The detected anomalies are then converted into human readable messages [13], which can be used to create a dedicated optimization measure [5]. Different anomaly detection algorithms can be used [14], whereas we identified the adjusted approach based on the Z-Score of the data points as most robust for different use cases in our context. In this adjusted approach, the thresholds for anomaly detection are not constant, but are defined dynamically dependent on the number of samples, their statistical distribution as well as user-defined settings (patent filed [15]). The approach was validated and adjusted based on real data and feedback from four large, international companies.

Figure 1 drafts ML recommendations from our research project and how both approaches could potentially be integrated with IS for product costing. The left side shows the similar parts search containing the list of similar materials for a hexagonal screw, which can be replaced by either the same material from another plant (PT1) or another material (M16). The right side shows results of plausibility checks. In addition to the anomalies identified, the potential cost impact of these anomalies on the product-cost estimate are also calculated and displayed.

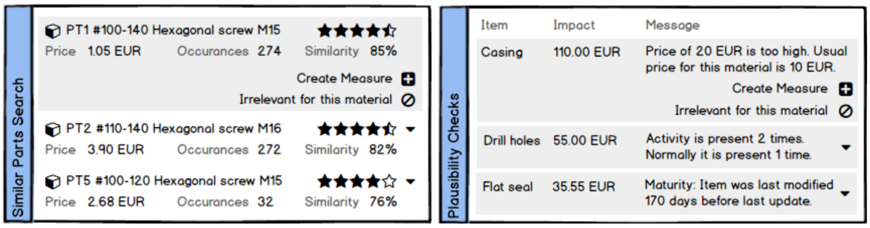


Figure 1. User interface mockups for the similar parts search and plausibility checks

2.2 Problem Case: Labeled Data

Both the similar parts search and plausibility checks can be realized with unsupervised [16] and supervised learning [17]. The latter requires a high quantity and quality of data (e.g., labels, master data, costing information, and bill of materials) to accurately train the ML model. As mentioned above, it is difficult to obtain the required large labeled datasets, given the need for extraordinary manual effort and the variety of potential errors in the costing structures [18]. Moreover, specific domain knowledge in product costing is required to accurately label the data [2], [9], [19].

To address those setbacks, which are quite common in ML [20], the presented approaches start with unsupervised learning using fuzzy search [11] as well as statistical and hierarchical algorithms [14]. Although these approaches are already a good support in improving the product-cost estimate quality, they require company-specific calibration and do not deal well with special use cases, which may result in false positives or missed errors. To combine the ML with the knowledge from the user without overwhelming the latter with additional efforts, we propose integrating the ML algorithms with our approach of MM [8]. We extend the approach with supervised learning to further improve the ML algorithms and thereby raise the quality of optimization recommendations (Figure 2).

With the initial ML model trained by unsupervised learning, we use inference to generate recommendations. The user now has the option to either mark the recommendations as irrelevant or, if it seems reasonable, initiate an optimization measure for further processing (Figure 1). All recommendations marked as irrelevant are flagged as false positives. With the initiation of an optimization measure, the respective recommendation is labeled implicitly as true positive and later labeled again after the measure has successfully been applied to the product [5].

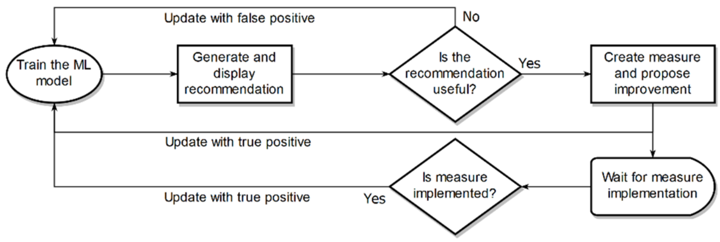


Figure 2. Measure management as an enabler of supervised learning

In sum, positive and negative feedback is used to retrain the models so future recommendations are more accurate (Figure 2). Thus, the user's feedback is reused without additional effort for labeling data. In that way, we can continuously improve learning results for both ML approaches and increase the accuracy of algorithms, even if optimization recommendations are not ultimately applied to the product.

3 Status Quo and Outlook

The initial unsupervised learning of the described optimization approaches has already been implemented in the context of our research partner's system SAP Product Lifecycle Costing (see [21]) to propose optimization potentials [12], [13]. To date, unsupervised learning has been performed with a set of real data representing four large, international companies in the discrete manufacturing industry, each with more than 10,000 employees. For them, plausibility checks detect anomalies and potential errors for up to 5% of all available line items in the provided costing structures. This error rate is surprisingly high, which can be explained on the one hand by undiscovered errors caused by manual editing of costing structures with up to 100,000 line items and on the other hand by the fact that the unsupervised approaches result in false positives in specific use cases. Therefore, we look forward to validating and further training of the ML models with the help of our presented approach.

With continuous improvement of the models and more profound results, further automation of the process is possible. For example, the most certain recommendations can be automatically implemented for entire product development projects with minimal manual effort. Even the creation of new costing structures based on identified typical costing structures and user-provided parameters is conceivable.

References

1. Stark, J.: Product Lifecycle Management: 21st Century Paradigm for Product Realisation, 3rd ed. Springer, London (2015)
2. Walter, M., Leyh, C.: Knocking on Industry's Door: Needs in Product-Cost Optimization in the Early Product Life Cycle Stages. In: Proceedings of the IEEE 19th Conference on Business Informatics (CBI 2017), pp. 330–338. Thessaloniki, Greece (2017)
3. Eigner, M., Stelzer, R.: Product Lifecycle Management: Ein Leitfaden für Product Development und Life Cycle Management. Springer, Heidelberg (2009)
4. Ehrlenspiel, K., Kiewert, A., Lindemann, U.: Cost-Efficient Design. Springer, Berlin (2007)
5. Walter, M., Leyh, C., Strahringer, S.: Knocking on Industry's Door: Needs in Product-Cost Optimization in the Early Product Life Cycle Stages. Complex Systems Informatics and Modeling Quarterly 13, 43–60 (2017)
6. Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77 (2008)
7. Peffers, K., Tuunanen, T., Niehaves, B.: Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. European Journal of Information Systems 27 (2), 129–139 (2018)
8. Walter, M., Leyh, C., Strahringer, S.: Toward Early Product Cost Optimization: Requirements for an Integrated Measure Management Approach. In: Proceedings of the Multikonferenz Wirtschaftsinformatik 2018 (MKWI 2018), pp. 2057–2068. Leuphana University Lüneburg, Lüneburg (2018)
9. Mörtl, M., Schmied, C.: Design for cost - a review of methods, tools and research directions. Journal of the Indian Institute of Science 95, 379–404 (2015)
10. Bell, J.: Machine Learning: Hands-On for Developers and Technical Professionals. Wiley, Indianapolis (2015)
11. SAP HANA Developer Guide: FUZZY Search, https://help.sap.com/viewer/691cb949c103 4198800afde3e5be6570/2.0.01/en-US/cc602780bb5710148aa2bf6cab3c015b.html (Accessed: 13.11.2018)
12. Heidenreich, M.: Unlock the Potential of Machine Learning to Get Better Cost Estimates in the Early Stage of the Product Lifecycle, https://blogs.sap.com/2017/10/04/unlock-the-potential-of-machine-learning-to-get-better-cost-estimates-in-the-early-stage-of-the-product-lifecycle/ (Accessed: 30.08.2018)
13. Vosough, Z., Vasyutynskyy, V.: Using Parallel Sets for Visualizing Results of Machine Learning Based Plausibility Checks in Product Costing. In: Proceedings of VisBIA 2018: Workshop on Visual Interfaces for Big Data Environments in Industrial Applications, pp. 4–11. Grosseto, Italy (2018)
14. Mehrotra, K. G., Mohan, C., Huang, H.: Anomaly Detection Principles and Algorithms. Springer, Cham (2017)
15. Wasiutinski, V.: Plausibility-Based Validation of Product Costing Structures. US Patent Filed (2018)
16. Hastie, T., Tibshirani, R., Friedman, J.: Unsupervised Learning. In: The Elements of Statistical Learning, pp. 485–585. Springer, New York (2009)
17. Caruana, R., Niculescu–Mizil, A.: An Empirical Comparison of Supervised Learning Algorithms. In: Proceedings of the 23rd International Conference on Machine Learning (ICML '06), pp. 161–168. New York (2006)
18. Settles, B.: Active Learning Literature Survey. Technical report 1648, Computer Sciences. University of Wisconsin-Madison (2009)

19. Rush, C., Roy, R.: Expert Judgement in Cost Estimating: Modelling the Reasoning Process. Concurrent Engineering 9, 271–284 (2001)
20. Zhu, X.: Semi-Supervised Learning Literature Survey. Technical report 1530, Computer Sciences. University of Wisconsin-Madison (2005)
21. Product Costing | SAP Product Lifecycle Costing, https://www.sap.com/products/product-lifecycle-costing.html (Accessed: 3.09.2018)

Bedeutung von Predictive Analytics für den theoretischen Erkenntnisgewinn in der IS-Forschung

Giacomo Welsch, Matthias Hauser, Frédéric Thiesse

Universität Würzburg, Wirtschaftswissenschaftliche Fakultät, Würzburg, Deutschland
{giacomo.welsch,matthias.hauser,frederic.thiesse}
@uni-wuerzburg.de

Abstract. Obgleich mehrere Forscher die Wichtigkeit von *Predictive Analytics* für den theoretischen Erkenntnisgewinn hervorgehoben haben, werden IS-Theorien der Verhaltensforschung in der Regel nicht auf ihre Vorhersagekraft getestet. Vor diesem Hintergrund soll in dem diesem Artikel zugrundeliegenden Forschungsprojekt am Beispiel der Technologieakzeptanzforschung aufgezeigt werden, dass der Einsatz von *Predictive Analytics* bei der Aufstellung von Theorien nicht nur wichtig ist, sondern sogar notwendig sein kann um die potenziellen Schwächen rein erklärender Modelle (insbesondere Vollständigkeit, Generalisierbarkeit und Praxisrelevanz) zu adressieren. Hierzu wird in diesem Artikel eine dreistufige Forschungsagenda zur Untersuchung und gegebenenfalls Modifizierung bestehender Modelle mithilfe von *Predictive Analytics* aufgezeigt.

Keywords: Technologieakzeptanzforschung, IS-Theorien, Predictive Analytics, Verhaltensforschung

1 Vorhersagemodelle in der IS-Forschung

Die Verhaltensforschung und die Design-Forschung gelten als die beiden Säulen der IS-Forschung [28]. Während sich Letztere in erster Linie mit dem Entwurf nützlicher Artefakte beschäftigt [4, 13], ist das Ziel der Verhaltensforschung die Aufdeckung von *Wahrheit* [19]. Theorien der Verhaltensforschung sollen hierbei insbesondere in der Lage sein, Verhalten einer bestimmten Zielgruppe (z.B. Kunden, Mitarbeiter oder Organisationen) gegenüber Artefakten zu erklären oder vorherzusagen [16].

Der Fokus dieser Arbeit liegt auf den Vorhersageeigenschaften von Modellen der Verhaltensforschung. Ausgangspunkt des Forschungsprojekts ist die Beobachtung, dass Theorien der Verhaltensforschung in der Regel nicht auf ihre Vorhersagekraft getestet werden [9, 23, 24]. Dies verwundert, da sich einflussreiche Forscher [12, 24] intensiv mit der Vorhersageeigenschaft von IS-Theorien auseinandergesetzt haben und von den *Grand Theories* sowohl Erklärungs- als auch Prognosefähigkeiten erwarten. Stattdessen werden Theorien der Verhaltensforschung in der Regel wie folgt entwickelt: Forscher stellen auf Basis theoretischer Überlegungen und bisheriger Forschungsergebnisse Hypothesen kausaler Zusammenhänge auf, die mithilfe einer Stichprobe an Daten überprüft werden. Werden die vermuteten Zusammenhänge im

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Zuge dessen als ausreichend stark und signifikant bewertet, gilt die jeweilige Theorie als bestätigt [5, 23]. Beispiele hierfür sind TAM [7] und UTAUT [26, 27].

Generell weist die IS-Forschung einen Mangel an Vorhersagemodellen auf [24]. Mehrere Forscher haben daher zu mehr Einsatz von *Predictive Analytics* aufgerufen, insbesondere weil sich mithilfe solcher Modelle theoretische Erkenntnisse gewinnen lassen können [1, 24]. In dem diesem Artikel zugrundeliegenden Projekt soll am Beispiel der Technologieakzeptanzforschung gezeigt werden, dass der Einsatz prädiktiver Verfahren bei der Aufstellung von Theorien notwendig ist. Darauf aufbauend soll ein Ansatz entwickelt werden, der aufzeigt, wie prädiktive Verfahren bei der Modifikation und Entwicklung von Theorien eingesetzt werden können.

2 Erklärungen und Vorhersagen in theoretischen Modellen

Theorien der Verhaltensforschung sollen Phänomene entweder beschreiben, erklären und/oder vorhersagen. Hierbei vernachlässigen rein beschreibende und rein prädiktive Theorien – im Gegensatz zu erklärenden Theorien – die Aufdeckung der latenten Mechanismen, die der Entstehung eines Phänomens zugrunde liegen [12]. Diese werden von dem Statistiker Breiman [5] als *Natur* beschrieben. Da rein beschreibende Theorien die einfachste Form einer Theorie sind und zum Einsatz kommen, wenn noch wenig über das zu betrachtende Phänomen bekannt ist [10], werden diese in diesem Artikel nicht näher betrachtet. Im Fokus liegen folglich Theorietypen zur (i) Erklärung, (ii) Vorhersage und (iii) Erklärung und Vorhersage (Abbildung 1).

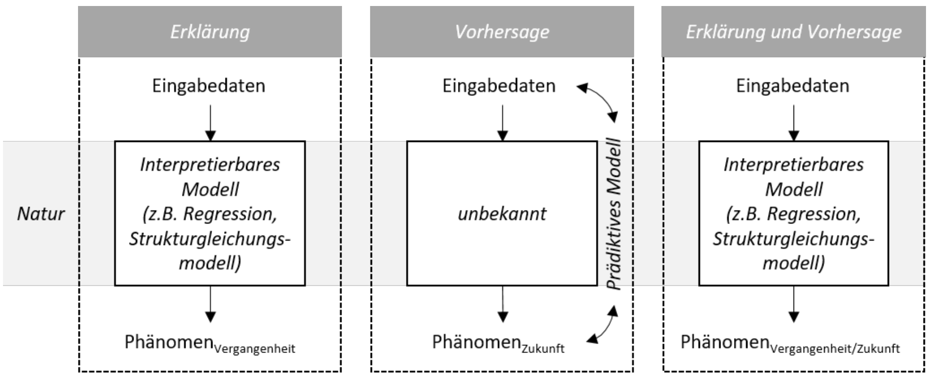


Abbildung 1: Theorietypen (in Anlehnung an Breiman [5] und Gregor [12])

Aus philosophischer Sicht unterscheiden sich Erklärungen und Vorhersagen dadurch, dass sich Erstere auf die Vergangenheit und Letztere auf die Zukunft beziehen [14, 20, 21]. Für die statistische Modellierung bedeutet dies folgendes:

- Bei Theorien zur reinen Erklärung werden für die Aufstellung eines Modells alle zu einem bestimmten Zeitpunkt verfügbaren Daten zu dem betrachteten Phänomen berücksichtigt, wobei insbesondere überprüft wird, ob die a priori vermuteten

Beziehungen in den Daten bestätigt werden können [5, 11, 23, 24]. Rein erklärende Modelle sind gut interpretierbar, liefern aber nicht per se gute Vorhersagen [5, 24].

- Rein prädiktive Modelle sind typischerweise das Ergebnis algorithmischer Verfahren wie beispielsweise Künstlicher Neuronaler Netze [5, 24]. Dabei sind die vorherzusagenden (Test-) Daten des Phänomens während der Aufstellung des Modells unbekannt. Es wird nur mit einem Teil der verfügbaren (Trainings-) Daten aufgestellt. Die Modellgüte wird bestimmt, indem überprüft wird, wie gut es die unbekannten Daten vorhersagen kann. Rein prädiktive Modelle sind in der Regel schwierig bis gar nicht interpretierbar, liefern jedoch sehr gute Vorhersagen [5].
- Für die Aufstellung von Modellen, die sowohl interpretierbar sind als auch gute Vorhersagen liefern, existiert kein eindeutiger Modellierungsansatz – entweder wird ein erklärendes Modell auf seine Vorhersagekraft getestet und gegebenenfalls modifiziert, um diese zu verbessern oder die Interpretierbarkeit eines prädiktiven Modells wird (zumeist auf Kosten der Vorhersagekraft) mithilfe entsprechender Methoden und Prädiktoren erhöht [24].

Laut Gregor [12] gehören *Grand Theories* wie TAM zu den Theorien der Vorhersage und Erklärung – dennoch werden sie im Rahmen ihrer Modellierung und Evaluation in der Regel wie rein erklärende Theorien behandelt [7, 24, 26, 27]. Auch wenn sich TAM und UTAUT in einer Vielzahl von Studien bewährt haben und im Rahmen dieser Arbeit nicht an ihrer Qualität gezweifelt werden soll, birgt rein erklärende Modellierung ohne prädiktive Evaluation Gefahren bezüglich der folgenden Kriterien:

- **Vollständigkeit.** Da der Fokus rein erklärender Modellierung auf der Bestätigung von Hypothesen liegt, besteht die Gefahr, dass die vollständige Aufdeckung der Mechanismen, die ein Phänomen verursachen, in den Hintergrund gerät. Fiedler [11] argumentiert daher, dass Studien, die den Fokus von der Bestätigung einzelner Zusammenhänge auf die vollständige Beschreibung eines Phänomens legen, ebenfalls notwendig sind. Solche Studien bezeichnet er als *phänomengetrieben*. Prädiktive, datengetriebene Studien legen den Fokus auf die Vorhersage von Phänomenen [24] und sind daher ideale Kandidaten für derartige Studien.
- **Generalisierbarkeit.** *Grand Theories* sollen weitestgehend ungebunden an Raum und Zeit sein [3]. Dies gilt in der Verhaltensforschung zwar nicht gleichermaßen wie in der naturwissenschaftlichen Forschung [2, 6, 18], eine möglichst generalisierbare Lösung sollte dennoch das Ziel sein [12]. Die größte Gefahr für die Generalisierbarkeit ist die *Überanpassung* (*Overfitting*) [23] – rein erklärende Modelle können auf die Daten ihrer Stichprobe überangepasst sein [17].
- **Praxisrelevanz.** Technologieakzeptanzmodelle sollen praxisrelevant sein – so soll UTAUT2 der Konsumgüterindustrie helfen, Technologien besser zu vermarkten [27]. Dazu ist es notwendig, dass die Erkenntnisse der Theorie auf die Zukunft übertragbar sind – hängt die Nutzungsabsicht in der Zukunft von anderen Faktoren ab als in der Vergangenheit, liefern die theoretischen Erkenntnisse keinen praktischen Mehrwert. Mithilfe prädiktiver Methoden kann daher die Praxisrelevanz von Theorien überprüft werden [24], mit rein erklärender Modellierung nicht zwingend.

3 Forschungsagenda

Im Rahmen dieses Projekts sollen Theorien der Technologieakzeptanzforschung auf ihre Vorhersagekraft überprüft und gegebenenfalls modifiziert werden, wobei die oben genannten Gefahren rein erklärender Modellierung vermieden werden sollen. Zu diesem Zweck sollen drei Studien durchgeführt werden:

1. **Vorhersagbarkeit der Nutzungsabsicht.** Mit *Predictive Analytics* kann die Vorhersagbarkeit von Phänomenen quantifiziert werden [8]. Lässt sich ein Phänomen mit einem prädiktiven Modell wesentlich besser vorhersagen als mit der zugrundeliegenden Theorie, spiegelt die Theorie die *Natur* noch nicht hinreichend wider [24]. Dabei kann sich ein prädiktives zu einem bestehenden theoretischen Modell dahingehend unterscheiden, dass es über weitere/alternative Prädiktoren verfügt und/oder mit anderen Methoden entwickelt wird. In einer explorativen Vorstudie zu diesem Forschungsprojekt konnte bereits eine entsprechende Prognosediskrepanz der Nutzungsabsicht zu einer etablierten Theorie der Technologieakzeptanzforschung mit einer kleinen Umfragestichprobe (N=249) festgestellt werden. Die verbesserte Vorhersage wurde dabei durch die Erweiterung und Kompression der von der Theorie suggerierten Prädiktoren unter Einsatz von *PLSpredict* und Kreuzvalidierung erzielt [25]. Da jedoch insbesondere die Analyse großer Datenmengen das Potenzial bietet, theoretische Erkenntnisse zu gewinnen [1, 24], soll in einer Folgestudie eine weit größere Stichprobe, die in Kooperation mit einem Praxispartner erhoben wird, mit maschinellen Lernverfahren analysiert werden. Ziel hierbei ist die Aufdeckung und Analyse einer substanziellen Diskrepanz zwischen den Vorhersagen des prädiktiven Modells und denen der Theorie sowie die Identifikation potenzieller neuer und alternativer Prädiktoren.
2. **Identifikation des *wahren* Modells.** Mithilfe einer Monte-Carlo-Simulation – ähnlich den Simulationen von Henseler et al. [15] und Rönkkö und Evermann [22] – sollen in einer zweiten Studie Daten eines *wahren* Strukturgleichungsmodells simuliert werden. Mit diesen Daten werden in einem ersten Schritt Strukturgleichungsmodelle aufgestellt, die das *wahre* Modell approximieren sollen. Anschließend werden diese Modelle jeweils auf ihre (i) erklärende und (ii) prädiktive Stärke überprüft. Hierbei werden im ersten Fall alle simulierten Daten herangezogen; im zweiten Fall wird der Datensatz in Trainings- und Testdaten partitioniert. Ziel der Studie ist zu zeigen, dass durch rein erklärende Evaluationen Modelle, die nicht dem *wahren* Modell entsprechen, bestätigt werden können. Darüber hinaus soll gezeigt werden, dass jenes Modell, das dem *wahren* Modell entspricht, bessere Vorhersagen liefert als die übrigen Modelle. Für die Berechnung der Vorhersagen werden verschiedene Ansätze der Strukturgleichungsmodellierung herangezogen.
3. **Vorgehen zur Modifikation von Theorien.** Ziel der dritten Studie ist mithilfe der Erkenntnisse der ersten beiden Studien einen Modellierungsansatz für die Aufstellung und Modifikation von Modellen zur Vorhersage und Erklärung zu entwickeln.

Literaturverzeichnis

1. Agarwal, R., Dhar, V.: Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. Information Systems Research 25, 443–448 (2014)
2. Audi, R. (ed.): The Cambridge Dictionary of Philosophy (2nd ed.). Cambridge University Press, Cambridge (1999)
3. Bacharach, S.B.: Organizational Theories: Some Criteria for Evaluation. The Academy of Management Review 14, 496–515 (1989)
4. Baskerville, R., Baiyere, A., Gergor, S., Hevner, A., Rossi, M.: Design Science Research Contributions: Finding a Balance between Artifact and Theory. Journal of the Association for Information Systems 19, 358–376 (2018)
5. Breiman, L.: Statistical Modeling: The Two Cultures. Statistical Science 16, 199–215 (2001)
6. Cook, T.D., Campbell, D.T.: Quasi-experimentation. Design & Analysis Issues for Field Settings. Houghton Mifflin, Boston (1979)
7. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35, 982–1003 (1989)
8. Ehrenberg, A.S.C., Bound, J.A.: Predictability and Prediction. Journal of the Royal Statistical Society Series A 156, 167–206 (1993)
9. Evermann, J. and Tate, M.: Comparing Out-of-Sample Predictive Ability of PLS, Covariance, and Regression Models. In: 35th International Conference on Information Systems, pp. 1–18. Association for Information Systems (AIS), Atlanta (2014)
10. Fawcett, J., Downs, F.S.: The Relationship of Theory and Research. Appleton-Century-Crofts, Norwalk (1986)
11. Fiedler, K.: What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing. Perspectives on Psychological Science 12, 46–61 (2017)
12. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
13. Gregor, S., Jones, D.: The Anatomy of a Design Theory. Journal of the Assocation for Information Systems 8, 312–335 (2007)
14. Hempel, C.G., Oppenheim, P.: Studies in the Logic of Explanation. Philosophy of Science 15, 135–175 (1948)
15. Henseler, J., Dijkstra, T.K., Sarstedt, M., Ringle, C.M., Diamantopoulos, A., Straub, D.W., Ketchen, D.J., Hair, J.F., Hult, G.T.M., Calantone, R.J.: Common Beliefs and Reality About PLS. Organizational Research Methods 17, 182–209 (2014)
16. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–105 (2004)
17. Hitchcock, C., Sober, E.: Prediction versus Accommodation and the Risk of Overfitting. The British Journal for the Philosophy of Science 55, 1–34 (2004)
18. Hospers, J.: An Introduction to Philosophical Analysis. Prentice-Hall, Englewood Cliffs (1967)
19. March, S.T., Smith, G.F.: Design and Natural Science Research on Information Technology. Decision Support Systems 15, 251–266 (1995)
20. Popper, K.: Logik der Forschung. Zur Erkenntnistheorie der Modernen Naturwissenschaft. Springer, Vienna (1935)
21. Rescher, N.: On Prediction and Explanation. The British Journal for the Philosophy of Science 8, 281–290 (1958)
22. Rönkkö, M., Evermann, J.: A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. Organizational Research Methods 16, 425–448 (2013)

23. Shmueli, G.: To Explain or to Predict? Statistical Science 25, 289–310 (2010)
24. Shmueli, G., Koppius, O.R.: Predictive Analytics in Information Systems Research. MIS Quarterly 35, 553–572 (2011)
25. Shmueli, G., Ray, S., Velasquez Estrada, J.M., Chatla, S.: The Elephant in the Room: Predictive Performance of PLS Models. Journal of Business Research 69, 4552–4564 (2016)
26. Venkatesh, V., Morris, M.G., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27, 425–478 (2003)
27. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 36, 157–178 (2012)
28. Winter, R.: Design Science Research in Europe. European Journal of Information Systems 17, 470–475 (2008)

Track 6:

Digitale Transformation und Dienstleistungen

Track Chairs:

Prof. Dr. Daniel Beverungen
Universität Paderborn

Prof. Dr. Christiane Lehrer
Universität St. Gallen

Heuristic Theorizing in Software Development: Deriving Design Principles for Smart Glasses-based Systems

Lisa Berkemeier[1], Benedikt Zobel[1], Sebastian Werning[2], Jannis Vogel[1], Florian Remark[1], Ingmar Ickerott[2], and Oliver Thomas[1]

[1]University of Osnabrück Information Management and Information Systems, Germany
{firstname.lastname}@uni-osnabrueck.de
[2]Osnabrück University of Applied Sciences, Business Administration and Logistics Management, Germany
{s.werning, i.ickerott}@hs-osnabrueck.de

Abstract. Design knowledge on smart glasses-based systems is scarce. Utilizing literature analysis on software development publications, insights from the design and implementation of four smart glasses-based systems and expert interviews, we elicited 16 design principles to provide guidance in the development of future service support systems. Heuristic Theorizing is an abductive Design Science Research method, hitherto far too little known or little noticed, which was applied to conduct the research. We contribute to theory and practice with applicable design principles to support the development of smart glasses-based systems. Phenomena known to have an impact on the adoption of smart glasses are addressed by these design principles.

Keywords: Smart Glasses, Augmented Reality, Design Principles, Heuristic Theorizing, Instantiation

1 Introduction

Digital transformation is characterized by unavoidability, irreversibility, rapidity and uncertainty [1]. Thereby an essential driver for today's digitalization are mobile technologies [2]. In this domain Augmented Reality (AR) reached much attention in research and practice. Haborth [3] discovered that a dominant technical research focus exists and inevitably studies regarding user behavior are underrepresented. However, a holistic knowledge is essential for the adoption of new technologies [4].

One technology to take into consideration are smart glasses. Smart glasses-based systems are identified as an appropriate solution to assist services with intense information-needs [5], [6]. Based on three years of research and a certain amount of smart glasses-based prototypes we define the corresponding concept as a mobile eyewear having a display and further sensors installed, in sum called smart glasses, which proactively provides process information and instructions to the user or allows the user to collaborate with integrated IT-systems (Human-Computer Interaction) or other process participants. For example, by using integrated cameras the user is able to scan bar codes or to take pictures and videos while executing assigned tasks. As a

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

further benefit, a valid smart glasses-based system allows a hands-free interaction and therefore releases further efficiency. The underlying technology Augmented Reality is a generic technical concept, which describes displays that enhance the perceived real world by digital and virtual objects [7]. Smart glasses-based systems are discussed in particular to cope with the variety and wide range of service tasks [8].

New technical features exist that (amongst others) can track the user and take photos unnoticedly. Therefore, privacy is a vital issue not only in terms of compliance, but for user's trust and perceived privacy. Potential users refrain from the adoption of smart glasses as they criticize privacy risks as well as ergonomic and social consequences following the use of smart glasses [9]. To ensure the adoption of this promising support technology, iterative adjustments are required to design and implement smart glasses-based systems to encounter existing adoption barriers [10]. A research gap exists thereby because those systems are still under ongoing investigation and have not been fully implemented in practice to date. Due to a small number of cases, the factors and phenomena influencing the adoption and diffusion of smart glasses are not fully understood [11]. Hence, the research focus of this paper is embedded in a design science research (DSR) project that deals with the global research question:

(RQ) *How to design smart glasses-based systems for industrial service scenarios aiming to positively influence adoption and diffusion?*

During a three-year consortium research, we investigated the design and development of smart glasses-based systems. Uncertainties arose in terms of implementation, usability, privacy, acceptance, and project management resulting from the aforementioned research gap. Heuristic can be understood as an approach to handle situations where limited time and knowledge are involved in achieving with a high probability suitable solutions. This is in contrast with the imagination of unlimited rational thinking and resources to find out the optimum [12]. Marewski et al. [13] figured out that heuristics, i.e. simple cognitive approaches, can be in favor when uncertainty is involved due to omitting insignificant details. Heuristic Theorizing [14] is applied in this context to structure the problem space and generate regarding solution components in an iterative approach. Then, we combined our justificatory knowledge of smart glasses-based system development with design principles from IS literature.

Our presentation of research is structured as follows: To begin with, the theoretical background of smart glasses, their implications for service support systems and research projects, which applied Heuristic Theorizing are presented (cf. section 2). Afterwards, we introduce Heuristic Theorizing as our research approach and apply it to the initial research questions (cf. section 3). This is followed by detailed explanations of our execution (cf. section 4). Then we describe the derived design principles (cf. section 5) and state hardware guidelines (cf. section 6). We conclude with discussion and further research need statements (cf. section 7).

2 Related Work

Smart devices appear in everyday objects that become omnipresent and allow interconnection to a smart environment. Popular manifestations of smart devices are smart glasses [11]. The main advantage of smart glasses are the hands-free navigation and context-sensitive information provision [15], [16]. Smart glasses are in particular applicable to information-intensive and bi-manual tasks. However, potential users criticize privacy risks as well as ergonomic and social consequences following the use of smart glasses [9]. In IS research, several models and theories are applied to explain and measure the technology acceptance. Due to a small number of cases, the determinants for the acceptance of smart glasses are still not fully understood [11]. First articles in the IS domain present prototypical implementations of smart glasses-based service support systems and describe their individual and domain-specific design process (cf. [8], [17–22]). These researchers built on general guidelines for human-computer interaction (cf. [23–25]) or design guidelines provided by manufacturing companies such as Google [26] and Sony [27]. A few approaches aim to derive design principles for smart glasses-based service support systems, but again these are limited to individual use cases and domains [18], [28].

However, design guidelines and principles in human-computer interaction for an application must be extended and adapted to specific hardware Design knowledge on smart glasses-based service support systems is scarce and limited in their generalizability [cf. 28]. Design guidelines provided by the manufacturers are specific to individual models and focus on UI design [26], [27]. As this disruptive technology is part of a sociotechnical system, principles regarding the interaction, information provision and implementation in an existing system landscape are necessary. Design Principles convey prescriptive design knowledge provided by DSR [29]. This prescriptive knowledge results from the generalization of experiences and knowledge from individual implementations. In the course of research, we applied a heuristic approach to derive such design principles for smart glasses-based systems, because heuristic thinking is useful in design situations with high uncertainties [30]. For instance, Lienhard and Legner [31] used Heuristic Theorizing in a research project to evolve design principles for mobile health apps.

3 Research Approach

Heuristic Theorizing [14] was applied as a research strategy during an iterative system development grounded in a three-year consortium research. This DSR method is characterized by the use of heuristics to design a satisficing solution for an unstructured problem space. The design and implementation of smart glasses-based systems is such an unstructured problem space, because of the lack of design knowledge. Therefore, the entry point is the research gap for the design smart glasses-based systems. This problem space was structured continuously with reoccurring “problem structuring heuristics”. The findings are concretized and translated into action with the use of “design heuristics”. Results from heuristics are extracted through “heuristic synthesis”. The applied heuristics served to build a satisficing solution in the form of (i) four prototypes and (ii) 16 design principles to design and develop smart glasses-based systems. The alternating phases of heuristic synthesis and problem structuring/ design heuristics supported the course of research with dynamic problem-solving approaches.

Figure 1 states the methods applied as heuristics and the main artifacts that result from the synthesis. The sequence of steps that we passed through in order to derive design principles for smart glasses-based systems by using Heuristic Theorizing is traced with numbers in ascending order from (1) to (17). The designed principles are deduced from scarce design knowledge in IS literature [32], experiences made in the design and instantiation [33] of smart glasses-based systems and expert interviews [34] with AR software engineers.

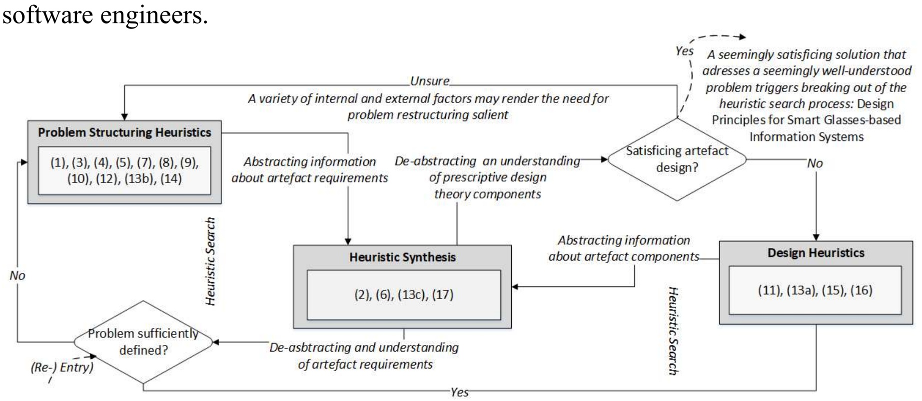


Figure 1. Applied process of Heuristic Theorizing

4 Executing Heuristic Theorizing

Overall, we applied nine different heuristics during research. To conduct the heuristic search, we used established *design science research methods and patterns* as published by Vaishnavi and Kuechler [35]. Initially, we *decomposed the research problem (1)* to identify the essential aspects of the research questions and to form the central problem class "design know-how for smart glasses-based systems" *(2)*. With an *interdisciplinary problem extrapolation (3),* we searched for existing solutions within other scientific disciplines such as law and psychology as well as solutions from practice. We found prototypes and studies from the field of logistics [36], manufacturing [37] and technical customer services [8], as well as design guides for user interfaces of AR or smart glasses [26], [27], [38]. Relevant knowledge was identified through *familiarization with the research community (4)* and the analysis of the *industry and practice awareness (5)* on smart glasses-based systems. The design then was enriched during the prototyping with insights from practice. We based our prototypes' design on already *published design solutions (6)*, both from research and from smart glasses producers [8], [26], [27], [39].

For the instantiation of our concept, we *identified a research domain (7)* and focused on a use case in logistics services as it inherits a linking function in the supply chain and provides numerous services, such as value-added services, which are highly intertwined with other industries. As we identified our research domain, we examined the *industry and practice awareness (8)* for smart glasses-based solutions in this specific sector again and collected requirements, challenges and potentials by conducting *expert interviews (9)* and a *systematic literature review (10)*. These expert interviews were conducted in the form of eight focus groups (two per prototype) consisting of three researchers (two from IS, one from logistics), two domain experts from logistics and two implementers of logistics systems. During these sessions, we discussed potential use cases for smart glasses in logistics and derived requirements for the prototyping. For an instantiation, the selection of the logistics domain is suitable as their workflows are characterized as information intense services with high flexibility [40]. Furthermore, smart glasses-based systems offer great potentials to improve logistic services. Potential use cases within the standard business processes in intralogistics have already been analyzed [41].

Software development activities, supported by design drafts *sketching possible solutions (11)*, were validated in four *expert interviews (12)* from the same group as before. Including the resulting feedback, four applications for smart glasses were developed through *iterative prototyping (13.a)*. In a cyclical attempt, we evaluated the applications using *experimentation and exploration (13.b)*. In a first step, students tested the applications and gave feedback. In a second step, the applications were tested with employees from logistics companies at real workstations. The feedback was collected in particular through discussions with the scientists involved and by applying the thinking aloud method. Four *smart glasses-based systems* for logistics support resulted from this approach *(13.c)*.

The four systems provide functionality for four different use cases. With system A, logistics workers can document damages of incoming goods by making use of the built-in camera. System B guides employees through the assembly processes of value-added services. System C provides a documentation method through a checklist-based approach for quality assurance, replacing paper-based lists. With system D, users can record processes while executing the respective tasks at the same time, which are then transformed into standardized process models automatically. While system A was developed for the Vuzix M100 smart glasses, the other three systems were built for its successor, the Vuzix M300. All four systems have been tested and evaluated by domain experts in real application cases. Major problems we identified in the course of research are the system's acceptance, usability, the ergonomic design, privacy and safety. We adapted the technical and functional design aspects regarding a modular architecture, measures to address acceptance and privacy compliance and included further aspects of the user interface and interaction patterns for smart glasses-based design. Unfortunately, adjustments for an improvement of the ergonomic design were limited due to the hardware. An example for the user interface and provided functionality of such system can be seen in Figure 2 (cf. section 5).

Investigating generalizability (14) of our design knowledge, we aggregated the similarities in problems and solutions among the four prototypes together with the participating developers to deduce the first set of *design principles (15)*. For evaluation this set was retrospectively discussed and was further supplemented in *interviews (16)*. One interview was conducted with the two implementation experts from the earlier interviews and one was conducted with two AR experts from a so far uninvolved software development company. The concepts and lessons learned from the development and implementation of the four systems were *embedded (17)* with the insights from literature and practice The three researchers mentioned have independently carried out a qualitative content analysis [42] and compared and integrated the results afterwards to develop a set of 16 design principles for smart glasses-based systems. Table 1 (cf. Appendix) states from which sources each design principle originates. In contrast to Niemöller et al [8], the proposed design principles are not based on a single implementation but integrate findings from four prototypical implementations and design knowledge from IS literature.

5 Design Principles

Overall, we identified 16 design principles, which we structure into four superordinate categories based on the architectural pattern of Model-View-Controller [43]. The category Interaction Design represents the controller (C), the category Software Design and Architecture serves as model (M). As smart glasses pose unique challenges concerning the presentation of data, the view-component (V) was divided into the categories Information Provision and User Interface, to state *which* and respectively information should be presented. Figure 2 states a summary of the proposed design principles. The design principles convey descriptive knowledge gained from the before mentioned DSR project and design knowledge from literature (cf. Table 1). This manifold design knowledge has been translated into action and materiality oriented [29] design principles. However, Information Provision deviates from this; these principles have been developed with a view to action as action-oriented design principles.

INTERACTION DESIGN (C)

(1) Simple and hands-free interaction utilizes value added by smart glasses. **(2)** Feedback from the system/ reciprocal and dynamic interaction generates steadyness for the user. **(3)** Take process and workplace safety requirements into account. **(4)** Using peripheral devices to optimize interoperability and the human computer-interface.

INFORMATION PROVISION (V)

(5) Focus on the essentials. **(6)** Match the user's qualification.
(7) Provide continuous navigation and orientation.

USER INTERFACE (V)

(8) Design a simple and consistent layout to avoid complexity. **(9)** Focus perceptible user interface design to ensure usability. **(10)** Offer customizable interfaces to respect individual user requirements.

SOFTWARE ARCHITECTURE AND DESIGN (M)

(11) Privacy is a design goal. **(12)** Use of Random Access Memory (RAM) economically. **(13)** Choose software components according to usage context. **(14)** Modular software design allows a customizable range of functions. **(15)** Scalable software architecture enables the future expansion of such young technology. **(16)** Use external sensors and interfaces to collaborate with others users and the overall.

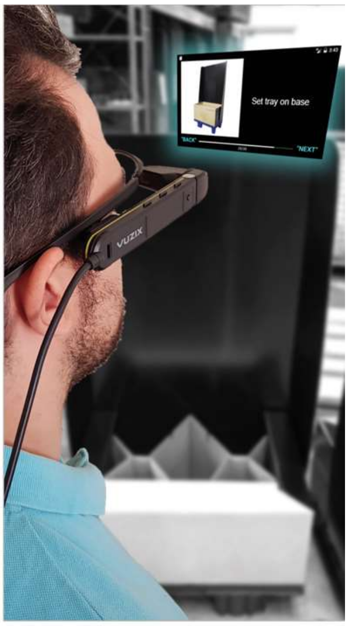


Figure 2. Design principles attributed to the model-view-controller paradigm

5.1 Interaction Design (C)

(1) Simple and hands-free interaction utilizes value added by smart glasses. To keep both hands free for bimanual tasks, e.g. picking of goods, voice control is favorable [21]. However in louder conditions, voice control can be distracted by noise or other voices. Alternative solutions such as gesture control or external control buttons should therefore also be provided.

(2) Feedback from the system/ reciprocal and dynamic interaction generates steadiness for the user. The feedback given by the smart glasses is important for the user's orientation among his tasks. To not lose track of the workflow, the user has to be informed about his handling errors and successful inputs [18], [21]. If voice control is deployed, the status of speech recognition should be displayed and an acoustic feedback could be implemented. We experienced confusion and mishandling if the user does not receive feedback on the status of data processing. If virtual objects are displayed in a smart glasses-based system with spatial localization, direction indicators should be implemented to guide the user to objects, which are placed outside his current field of view.

(3) Take process and workplace safety requirements into account. Smart glasses are deployed in industrial settings with potential dangerous conditions and processes. To safe the user from imminent danger, giving input during safety warnings should be disallowed. Regarding binocular devices, a half display mode, e.g. for forklift drivers, to ensure an unaffected perception of reality combined with larger font sizes during movement activities seem to increase the safety at work even further.

(4) Using peripheral devices to optimizes interoperability and the human-computer interface. Smart glasses as a standalone solution have limited resources. The connection of smart glasses with peripheral devices (e.g. telemetry sensors) increases the application potential and, for instance, enables smart services [8]. The possibility of a wireless connection to the internet makes online communication (e.g. video chat) possible [28]. Interoperability with other devices or the interaction with machines that are part of the user's working place can enhance the support of complex or collaborative tasks.

5.2 Information Provision (V)

(5) Focus on the essentials. Minimalistic and context-adaptive information provision and one main screen with crucial information about the task to fulfill, prevent the occurrence of information overload and cognitive stress [8]. Additional information of the displayed process should be attached and split into multiple screens if it is to complex or too much content [8]. To reduce confusion and to not lose track of the workflow, hierarchies should be kept flat, messages should be short and concise [27]. Complex structures also have to be avoided due to limited legibility and cognitive stress. Instead of this, physical objects seem suitable [38]. Finally, the navigation depth should be as small as possible to prevent information overflow and cognitive stress [8].

(6) Match the user's qualification. An information provision that matches the user's qualification leads to a higher acceptance and a better user experience. For example, novices might require additional information, while this additional information is redundant for experts. Moreover, the full workflow should be represented by one smart glasses device. This prevents media disruption and improves the acceptance and user experience [8].

(7) Provide continuous navigation and orientation. Avoiding the unexpected, e.g. sending content too frequently and at unexpected times, helps to increase acceptance and usability [26]. Providing a progress bar to keep track of information and always returning to the last shown step are also supporting acceptance and usability [8]. Recognizable icons guiding the user and let him know immediately about the meaning of the corresponding message. To allow direct feedback to one step is also improving the usability [8].

5.3 User Interface (V)

(8) Design a simple and consistent layout to avoid complexity. Complex and dynamic layouts impede the usage of smart glasses. The following features decrease such complexity: Virtual screens and look-around displays should have no more than three colors maximum [21], [26], [44]. The Information should be placed in the (lower) center of the screen [21], [44], navigation bars at the bottom [21], [26]. Intuitive elements, such as icons, find their place in the corners of the display [26]. A dark background for look-around displays or to display 2D objects is desirable. An adaptable mode for different usage contexts (e.g. day and night mode) is useful.

(9) Focus perceptible user interface design to ensure usability. To deliver a good perceptibility, it is important to use an appropriate font size and type [27]. Recognizable icons enhance an intuitive use and reduce complexity. In regards to AR glasses with spatial localization, depth cues for virtual objects and a well-suited field of view increase immersion and are crucial for a comfortable head movement [38]. Finally, real world objects should not be covered by virtual content. Digital objects should be placed in open spaces [38].

(10) Offer customizable interfaces to respect individual user requirements. The user should be allowed to adapt the colors and positioning of elements to match their individual requirements and information needs. As a consequence, stronger involvement of the user results and seem to lead to higher acceptance [28].

5.4 Software Design and Architecture (M)

(11) Privacy is a design goal. Security breaches and client insecurity, privacy risks as well as General Data Protection Regulation (GDPR) compliance are common problems with smart glasses that should be avoided. As a result privacy by design and default are central rules to be followed for smart glasses-based systems. Therefore, latest data security measures have to be taken [45]. Moreover, a transparent system architecture as well as a transparent data collection and storage increase the trust in the system and in doing so the users acceptance [27], [45].

(12) Use Random Access Memory (RAM) economically. The main goal of all measures regarding the economical use of RAM deals with a lack of battery power, heat generation, and speed reduction as well as stability problems. These issues are caused by still constrained hardware setups of latest smart glasses devices. The downscaling of images and the avoidance of unnecessary elements and activities reduce the extent of RAM use and therefore increase the performance as well as user experience of the smart glasses-based application.

(13) Choose software components according to the usage context. To meet the device properties and the system requirements, deploying a sufficient framework is important to meet the device properties and the system requirements. One the one hand significant factors for the choice of an implementation framework, are hardware, the complexity of the task, support and scalability. On the other hand, solution components may vary with the usage context and therefore require a certain level of flexibility. For example, in most cases integrated voice commands like a "Next" voice statement are sufficient to complete a task rather than implementing a cloud-based speech-to-text service (even causing further issues with regards to DP 11). According to the usage context of the resulting support system, corresponding applications may use different functions of the system. For example, different options of voice control can be applied (command vs. speech-to-text) instead of single commands (e.g. "say what you see" principle. Also, connectivity, algorithm, implementation effort and the level of maturity are the deciding factors for the choice of voice control as some require an internet connection.

(14) Modular software design allows for a customizable range of functions. A modular design of system components enhances scalability and customization. Turning off unnecessary functionalities (modules) for specific use cases, receives several benefits regarding privacy compliance, the complexity of the system and the extension of RAM usage.

(15) Scalable software architecture enables the future expansion of such young technology. The system should be maintainable, expandable and scalable to meet an increasing demand and technical evolution.

(16) Use external sensors and interfaces to collaborate with others users and the overall system environments. Based on an increasing amount of sensors around the company landscape, mainly driven by Internet-of-Things concepts, there is a great potential to access process data, e.g. watching live machine utilization or being informed in case a new truck is arriving. Furthermore special peripheral devices like hand-mounted scanners utilizing additional process performance.

6 Hardware-specific Guidelines

During the course of research we additionally identified actionable guidelines for the choice and preparation of hardware. As these hardware guidelines occurred as a byproduct of our discussions and information generating methods, they were not integrated into the previously presented design principles. Nevertheless, they serve as important insights concerning the choice and handling of smart glasses.

Ensure wearing comfort. Regarding the Hardware Design, ensuring the wearing comfort of smart glasses is the first identified guideline. Poorly balanced as well as a too heavy weighted devices may lead to muscular pain and discomfort during a long working day [46]. A flexible position of the display supports wearing comfort by providing a better individual perceptibility [18]. The stable attachment of the smart glasses at the head further enhances the wearing comfort during movement. To meet individual demands, the suitability of the devices for spectacle wearers and mounting options for both eyes are necessary.

Choose hardware usable in industrial settings. The application of smart glasses in an industrial context leads to specific requirements. The consideration of industry standards is crucial with regard to the acceptance of smart glasses [8], [28]. Certain working contexts such as outdoor or soiled environments require waterproof and shockproof housing. Clip-on solutions are favorable for mounting on existent equipment, to be compliant with safety regulations in many industries.

Minimalistic Hardware Design. Complex and fragile hardware impedes the adoption of smart glasses [8], [28]. Consequently, the functionalities should match the usage context, and unnecessary functions should be avoided. The outsourcing of computational power as well as the battery, combined with high mobility are also important factors to minimize the hardware. A modular system would comply with the privacy by design paradigm [45], because the built-in sensors and functionalities are limited to the usage scenario.

Ensure excellent legibility. The legibility of smart glasses displays depends on technical as well as environmental conditions and has to suit the individual requirements of the user. Adjustable brightness, high resolution, and protection against solar radiation are crucial for excellent legibility of the display [28], [38].

7 Discussion, Conclusion, and Outlook

Discussion. Through applying heuristic theorizing, we introduced design principles for smart glasses-based systems. These design principles enrich previous research [6], as they focus on how software for smart glasses should be designed and developed. We contribute to the limited knowledge-base of smart glasses-based system engineering by providing results from the successful application of Heuristic Theorizing [14], an appropriate method for systems engineering which hitherto is not very known in research. We designed and instantiated four systems on two different hardware models. Hitherto, an evaluation has been carried out in the form of an interview with implementation experts. The effectiveness of the proposed design principles was not part of previously documented studies, and hence needs to be addressed in subsequent steps. The design principles for look through devices with spatial localization originate from literature and expert interviews. Further instantiations are required to extend the specific design knowledge within the range of those smart glasses.

Conclusion. In a DSR approach, we utilized design knowledge published by IS scholars and smart glasses-producers, the design and instantiation of a smart glasses-based system and expanded these findings with knowledge provided by augmented reality software engineers. The lack of long-time experience with smart systems demands the adaption and extension of best practices to meet the requirements emerging with the implementation in business settings and to derive design knowledge for future implementations. Heuristic Theorizing enables the integration of scientific insights to meet the requirements of innovative technology. As an open and creative process, it contributes new insight and is beneficial to the artefacts for practical use.

With our research, we contribute to theory with design principles for smart glasses-based systems. The aggregated design principles address the phenomena of acceptance and privacy, as crucial factors for the fail of smart glasses on the consumer market, as well as usability and ergonomic design. Furthermore, the paper presents a respective domain that is affected by the digital transformation and demonstrates with the application of Heuristic Theorizing an approach to handle digitalization research projects.

Outlook. Technology acceptance and usability of pervasive computing such as smart glasses is not understood, yet. Especially impediments in smart gasses adoption such as acceptance [47] require further understanding. Therefore, influence factors regarding those devices need to be examined and integrated into theories for technology acceptance [11]. To spark future research, the presented design principles can further be validated, expanded and refined through further case studies. A combination of our findings with further insights from the research community and particularly from interdisciplinary areas, e.g. from cognitive science, psychology and ergonomics, is planned to deepen the understanding of phenomena influencing the adoption of smart glasses-based systems.

8 Acknowledgements

This research study is based on the project GLASSHOUSE (grant number 02K14A090), funded by the German Federal Ministry of Education and Research (BMBF) within the Program "Innovations for Tomorrow's Production, Services and Work" and managed by the Project Management Agency Karlsruhe (PTKA).

References

1. Krcmar, H.: Charakteristika digitaler Transformation BT - Digitale Transformation: Fallbeispiele und Branchenanalysen. In: Oswald, G. and Krcmar, H. (eds.) Digitale Transformation. pp. 5–10. Springer Fachmedien Wiesbaden, Wiesbaden (2018).
2. Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N., Ahlemann, F.: Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. Bus. Inf. Syst. Eng. 59, 301–308 (2017).
3. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review. Twenty-third Am. Conf. Inf. Syst. 1–10 (2017).
4. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manage. Sci. 35, 982–1003 (1989).
5. Herterich, M.M., Peters, C., Uebernickel, F., Brenner, W., Neff, A.A.: Mobile Work Support for Field Service : A Literature Review and Directions for Future Research. Int. Conf. Wirtschaftsinformatik. 134–148 (2015).

6. Niemöller, C., Metzger, D., Fellmann, M., Özcan, D., Thomas, O.: Shaping the Future of Mobile Service Support Systems - Ex-Ante Evaluation of Smart Glasses in Technical Customer Service Processes. In: Mayr, H.C. and Pinzger, M. (eds.) Informatik von Menschen für Menschen (Informatik 2016). LNI 259, Klagenfurt, Österreich (2016).
7. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst. 77, 1321–1329 (1994).
8. Niemöller, C., Metzger, D., Thomas, O.: Design and Evaluation of a Smart-Glasses-based Service Support System. In: 13. International Conference on Wirtschaftsinformatik (WI). pp. 106–120 (2017).
9. Koelle, M., Kranz, M., Andreas, M.: Don't look at me that way ! – Understanding User Attitudes Towards Data Glasses Usage. In: Mobile HCI '15. pp. 362–372 (2015).
10. Hobert, S., Schumann, M.: Enabling the Adoption of Wearable Computers in Enterprises - Results of Analyzing Influencing Factors and Challenges in the Industrial Sector. Proc. 50th Hawaii Int. Conf. Syst. Sci. 4276–4285 (2017).
11. Segura, S., Thiesse, F.: Extending UTAUT2 to Explore Pervasive Information Systems. In: ECIS 2015 Completed Research Papers. p. Paper 154 (2015).
12. Todd, P.M., Gigerenzer, G.: Precis of Simple heuristics that make us smart. Behav. Brain Sci. 23, 727–780 (2000).
13. Marewski, J.N., Gaissmaier, W., Gigerenzer, G.: Good judgments do not require complex cognition. Cogn. Process. 11, 103–121 (2010).
14. Gregory, R.W., Muntermann, J.: Generating Design Theories Heuristic Theorizing : Proactively Generating Design Theories. Inf. Syst. Res. 25, 639–653 (2014).
15. Reif, R., Günthner, W.A., Schwerdtfeger, B., Klinker, G.: Pick-by-Vision Comes on Age: Evaluation of an Augmented Reality Supported Picking System in a Real Storage Environment. In: Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa. pp. 23–31. , New York, NY, USA (2009).
16. Zheng, X.S., Matos, P.: Eye-Wearable Technology for Machine Maintenance : Effects of Display Position and Hands-free Operation. CHI 2015, Crossings. 2125–2134 (2015).
17. Metzger, D., Niemöller, C., Jannaber, S., Berkemeier, L., Brenning, L., Thomas, O.: The Next Generation: Design and Implementation of a Smart Glasses-based Modelling System. Enterp. Model. Inf. Syst. Archit. - Int. J. Concept. Model. [accepted for publication] (2018).
18. Berkemeier, L., Werning, S., Zobel, B., Ickerott, I., Thomas, O.: Der Kunde als Dienstleister: Akzeptanz und Gebrauchstauglichkeit von Smart Glasses im Self-Service. HMD Prax. der Wirtschaftsinformatik. 54, 781–794 (2017).
19. Zobel, B., Berkemeier, L., Werning, S., Vogel, J., Ickerott, I., Thomas, O.: Towards a Modular Reference Architecture for Smart Glasses-based Systems in the Logistics Domain Smart Glasses for Logistics Services Functions From a Research Project to a Reference Architecture. In: 9th International Workshop on Enterprise Modeling and Information Systems Architectures (EMISA). p. Paper 17 (2018).
20. Broach, J., Hart, A., Griswold, M., Lai, J., Boyer, E.W., Skolnik, A.B., Chai, P.R.: Usability and Reliability of Smart Glasses for Secondary Triage During Mass Casualty Incidents. Proc Annu Hawaii Int Conf Syst Sci. 9, 1416–1422 (2018).
21. Schwantzer, S.: Konzeption und Implementierung eines Smart-Glasses-basierten Informationssystems für technische Dienstleistungen. Digit. der Aus- und Weiterbildung. 94–112 (2018).
22. Vogel, J., Jannaber, S., Zobel, B., Thomas, O.: Design and Development of a Process Modelling Environment for Business Process Utilization within Smart Glasses. In: EMISA. pp. 85–89 (2018).

23. Shitkova, M., Holler, J., Heide, T., Clever, N., Becker, J.: Towards Usability Guidelines for Mobile Websites and Applications. Wi. 1603–1617 (2015).
24. Benyon, D.: Designing Interactive Systems: A comprehensive guide to HCI, UX and interaction design, 3/E, (2014).
25. Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N.: Designing the User Interface: Strategies for Effective Human-Computer Interaction. Pearson (2016).
26. Google: Design for Glass, https://developers.google.com/glass/design/.
27. Sony Corporation: Design guidelines | Sony Developer World, https://developer.sony.com/develop/wearables/smarteyeglass-sdk/design-guidelines/.
28. Berkemeier, L., Menzel, L., Remark, F., Thomas, O.: Acceptance by Design: Towards an Acceptable Smart Glasses-based Information System based on the Example of Cycling Training. In: Multikonferenz Wirtschaftsinformatik (MKWI 2018). pp. 1027–1038 (2018).
29. Chandra, L., Seidel, S., Gregor, S.: Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions. Proc. Annu. Hawaii Int. Conf. Syst. Sci. 2015–March, 4039–4048 (2015).
30. Neth, H., Gigerenzer, G.: Heuristics: Tools for an Uncertain World. Emerg. Trends Soc. Behav. Sci. 1–18 (2015).
31. Lienhard, K.R., Legner, C.: Principles in the Design of Mobile Medical Apps: Guidance for Those who Care. In: Leimeister, J.M. and BrennerW. (eds.) Proceedings der 13. Tagung Wirtschaftsinformatik (WI 2017). pp. 1066–1080. , St. Gallen, Schweiz (2017).
32. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: European Conference on Information Systems. pp. 2206–2217. AISeL, Verona (2009).
33. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Types of Knowledge in Design Science Research. MIS Q. 37, 337–355 (2013).
34. Misoch, S.: Qualitative Interviews. De Gruyter Oldenbourg, Berlin (2015).
35. Vaishnavi, V.K., Kuechler, W.J.: Design Science Research Methods and Patterns: Innovating Information and Communication Technology. Auerbach Publications, Boston, MA, USA (2007).
36. Schwerdtfeger, B., Reif, R., Günthner, W.A., Klinker, G., Hamacher, D., Schega, L., Böckelmann, I., Doil, F., Tümler, J.: Pick-by-Vision: A first stress test. Sci. Technol. Proc. - IEEE 2009 Int. Symp. Mix. Augment. Reality, ISMAR 2009. 115–124 (2009).
37. Yew, A.W.W., Ong, S.K., Nee, A.Y.C.: Towards a griddable distributed manufacturing system with augmented reality interfaces. Robot. Comput. Integr. Manuf. 39, 43–55 (2016).
38. Meta Company: Design – Meta User Interface Guidelines | Overview | Augmented Reality | Meta Company, https://devcenter.metavision.com/design/user-interface-guidelines-overview.
39. Ok, A.E., Basoglu, N.A., Daim, T.: Exploring the design factors of smart glasses. In: 2015 Portland International Conference on Management of Engineering and Technology (PICMET). pp. 1657–1664. IEEE (2015).
40. Macrae, N.: Workplace flexibility: Past, present, and future. J. Labor Res. 10, 51–55 (1989).
41. Niemöller, C., Zobel, B., Berkemeier, L., Metzger, D., Werning, S., Adelmeyer, T., Ickerott, I., Thomas, O.: Sind Smart Glasses die Zukunft der Digitalisierung von Arbeitsprozessen? Explorative Fallstudien zukünftiger Einsatzszenarien in der Logistik. In: Leimeister, J.M. and Brenner, W. (eds.) International Conference on Wirtschaftsinformatik. , St. Gallen, Schweiz (2017).
42. Mayring, P.: Qualitative content analysis: theoretical foundation, basic procedures and software solution. , Klagenfurt (2014).

43. Krasner, G.E., Pope, S.T.: A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 System. J. Object Oriented Program. 1, 26–49 (1988).
44. Tanuma, K., Sato, T., Nomura, M., Nakanishi, M.: Comfortable Design of Task-Related Information Displayed Using Optical See-Through Head-Mounted Display. In: Salvendy, G. and Smith, M.J. (eds.) Human Interface and the Management of Information. pp. 418–429 (2011).
45. Berkemeier, L., McGuire, M.-R., Steinmann, S., Niemöller, C., Thomas, O.: Datenschutzrechtliche Anforderungen an Smart Glasses-basierende Informationssysteme in der Logistik. In: Informatik 2017. pp. 1037–1048. Lecture Notes in Informatics (LNI), Chemnitz (2017).
46. Theis, S., Mertens, A., Wille, M., Rasche, P., Alexander, T., Schlick, C.M.: Effects of data glasses on human workload and performance during assembly and disassembly tasks. In: Proceedings of the 19th Triennial Congress of the IEA. pp. 1–8. , Melbourne (2015).
47. Koelle, M., El Ali, A., Cobus, V., Heuten, W., Boll, S.C.J.: All About Acceptability?: Identifying Factors for the Adoption of Data Glasses. Proc. 2017 CHI Conf. Hum. Factors Comput. Syst. 295–300 (2017).

Appendix

Table 1. Design principles grounded in literature and design heuristics

Design Principle	Literature	Design Heuristics	Example
(1) Simple and hands-free interaction utilizes value added by smart glasses.	[8], [18], [21], [28]	Sketching Solution	For actual freehand use, only voice control comes into question. Gesture recognition or input using buttons temporarily require at least one hand.
(2) Feedback from the system/ reciprocal and dynamic interaction generates steadiness for the user.	[18]	Iterative Prototyping	Users were irritated when the system did not confirm a successful entry.
(3) Take process and workplace safety requirements into account.		Iterative Prototyping	Users handle chemicals and require instructions during the process.
(4) Using peripheral devices to optimizes interoperability and the human-computer interface.	[28]	Sketching Solution, Iterative Prototyping	Users faced problems to scan barcodes with the camera. Instead, a barcode scanner ring could be added to the system.
(5) Design a Simple and consistent layout to avoid complexity	[21], [26], [27], [8]	Sketching Solution, Iterative Prototyping	Strongly differing layouts and forms of presentation irritate the user.
(6) Focus perceptible user interface design to ensure usability.	[18], [21], [26], [27],	Embedding concepts and Techniques	Some manufacturers provide specific information for display on the devices, but these are not transferable to other products.

Design Principle	Literature	Design Heuristics	Example
(7) Offer customizable interface to respect individual user requirements.	[28]	Iterative Prototyping	During the evaluation of a prototype with different people, differing requirements were placed on the user interface.
(8) Focus on the essentials.	[8], [18]	Iterative Prototyping	Implementers with experience in app development for smartphones tend to design the user interface and the information to be displayed too complex for smart glasses-based systems.
(9) Match the user's qualification.	[8]	Sketching Solution	Different user groups require a differing degree of detailing and additional information for process support.
(10) Provide continuous navigation and orientation.	[8], [38]	Iterative Prototyping	A progress bar provides users with an overview of the process step in which they are.
(11) Privacy is a design goal.	[19], [28], [45]	Sketching Solution	The regulations of the GDRP must be complied with for industrial use. Privacy-by-design is a fundamental principle to be integrated into system development.
(12) Use of Random Access Memory (RAM) economically.		Iterative Prototyping	The application has a high latency time or the smart glasses have already heated up after a short period of use.
(13) Choose software components according to usage context.		Iterative Prototyping	For step-by-step instructions, a different language application and library is suitable than for speech-to-text scenarios.
(14) Modular software design allows a customizable range of functions	[19], [28]	Iterative Prototyping	Technical and functional modules of already existing prototypes could be integrated in the implementation of new use cases. Thus a registration mask or a video recording function of different prototypes has been realized with the same system module.
(15) Scalable software architecture enables the future expansion of such young technology.	[19]	Sketching Solution	So far, smart glasses have initially been used as prototypes. However, the corresponding solutions should be scalable in order to enable broad use in the future. So far
(16) Use external sensors and interfaces to collaborate with others users and the overall system environments		Sketching Solution	In order to develop smart services, external sensor data, e.g. from machines to be maintained, are required.

Mirroring E-service for Brick and Mortar Retail: An Assessment and Survey

Jan H. Betzing[1], Marco Niemann[1], Benjamin Barann[1],

Benedikt Hoffmeister[1], and Jörg Becker[1]

University of Münster, ERCIS, Münster, Germany,
{betzing,niemann,barann,hoffmeister,becker}@ercis.de

Abstract. The digital transformation increasingly impacts the competitive retail market structure in favor of e-commerce and digital business models, while many Brick and Mortar (BaM) retailers are struggling to meet customers' expectations. Supported by the customer adaption of e-commerce and digital technologies, this paper applies the lens of channel complementary theory to BaM. We examine, which e-service touchpoints from e-commerce can be transferred to the physical servicescape of BaM retail to complement customer journeys. Drawing from the dominant design theory, we first assess leading e-commerce solutions to identify dominant e-service touchpoints, which are then mirrored for their application in BaM retail. Second, we surveyed 250 shoppers to elicit the likeliness of use regarding these touchpoints. Our results provide a foundation for both academia and retail to advance the knowledge of relevant e-service touchpoints in BaM.

Keywords: Brick and Mortar Retail, Omni-Channel, Touchpoint, Dominant Design Theory, Channel Complementary Theory

1 Introduction

The digital transformation is an ongoing organizational change process that leverages technology-enabled innovation to optimize existing business processes and to digitally (re-)engineer business models [1, 2]. In retail, the digital transformation increasingly impacts the competitive market structure in favor of e-commerce and digital retail business models and also bolsters customers' expectations towards e-service offerings in BaM [3–5]. In particular, small and medium-sized enterprise (SME) Brick and Mortar (BaM) retailers, which comprise the majority of stores in Germany, face strong competition with pure online players and large retail chains that embrace technological innovations [6, 7]. Over the last years, e-commerce in Germany steadily realized double-digit growth rates [8, 9], while SME retail turnovers are projected to decline by up to 30 % until 2020 in many cities [10]. To cope with the digital transformation, large retail chains are heavily investing in channel management to become omni-channel retailers that are able to seamlessly interact with customers through any physical and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

digital channel expected by them [11, 12]. While selecting appropriate channels and technologies to adopt for this purpose is already challenging for large retail corporations [13, 14], it can appear like a Herculean task for SME BaM retailers [7]. However, BaM retailers that are unable to develop new (digital) strategies to address customers' expectations, are threatened to become "amazon-ed" [15]. When asking customers for their reasons to choose e-commerce over BaM retail, comparably lower prices to BaM retail (36 %,[8])—contrary to popular belief—are only one aspect mentioned, while a considerable number of customers also names comfort and ease of use (50 %, [8]), 24/7 availability (39 %, [8]), and low purchasing efforts as motivators [8, 16]. Hence, non-financial levers do exist in e-commerce that could turn out to be beneficial in the BaM context (e.g., the integration of 24/7 e-service to increase BaM store visibility).

Particular to e-commerce is the offering of *e-service touchpoints*. An e-service is understood as the application of digital competencies "through deeds, processes, and performances for the benefit of another entity or the entity itself" [17, p.26], where the "delivery is mediated by information technology" (IT) [18, p.341]. Literature differentiates touchpoints into instances and classes [19]. The former touchpoints describe moments of contact, i.e., any "instance of communication between a customer and a service provider" [20, p.846]. This contact belongs to a touchpoint class, which denotes "an abstract interaction interface to the customer" [19, p.4]. When using the term (e-service) touchpoint, we refer to a touchpoint class. Established technologies that provide access to these touchpoints in the physical servicescape of BaM retail are, for example, smartphone and in-store terminal applications (apps) [21, 22]. Against this background, it is fair to assume that *mirroring*—adapting e-service touchpoints, originally provided in e-commerce, to the physical servicescape—represents a good starting point for BaM retailers to transition towards an omni-channel environment.

Still, retailers can select from a wide variety of conceivable e-service touchpoints to offer as value propositions that potentially create value for and with the customer [17, 23, 24]. In particular, it remains unclear, whether customers in the physical servicescape of BaM retail are likely to use the same e-service offerings they are used to in e-commerce. Empirical studies on omni-channel retail are sparse [13], and BaM retailers are lacking guidance regarding the selection and mirroring of e-service touchpoints [22]. To shed light on this matter, our research question reads as follows: *What e-service touchpoints, enabled by smartphone and in-store terminal applications, should be mirrored to the physical servicescape of brick and mortar retail?*

We followed a two-step approach to answer this question. First, we identify a set of major touchpoints offered by leading e-commerce platforms, group them, and mirror them for their application in a traditional BaM retail servicescape. Second, we surveyed 250 shoppers to elicit the likeliness of use regarding these touchpoints.

The remainder of this paper unfolds as follows: Section 2 positions our study within the body of knowledge. Section 3 presents the set of mirrored e-service offerings, whereas Section 4 introduces our survey research method and the survey results. Section 5 discusses our results and relates them back to theory. We close with an overview of our contributions and an outlook on future research in Section 6.

2 Theoretical Background

Customers have become accustomed to an increasing level of digital touchpoints in all facets of their lives [25], and also expect e-service offerings in traditional BaM environments [12, 21]. In this context, the term hybrid customer interaction is used to describe a behavior, where customers are present simultaneously in both the physical and digital worlds [26]. For example, customers use their smartphones in-store as a "second screen" to compare prices, while interacting with a salesperson [27]. Phenomena such as "showrooming" [5] or using e-service touchpoints by other service providers during the store visit can be cannibalizing for BaM retailers.

On the other hand, hybrid customer interaction can also be leveraged by retailers, when they are able to provide own e-service touchpoints for customers to use in-store [28]. Our reasoning follows channel complementary theory [29], which explains that traditional and new media formats for implementing communicative functions may exist alongside instead of replacing each other. The theory suggests that the function is a more important consumption driver than the medium itself [29]. In our context, e-service touchpoints can be understood as such communicative functions. Some authors provide evidence of channel synergies in retail [30] and suggest a "bricks-and-clicks" approach, where insights from the BaM channel are integrated into an online channel [28]. However, the opposite direction—adapting knowledge from e-commerce to the physical BaM servicescape—has not received much attention. Further, most omni-channel initiatives focus on integrating *existing* digital and physical channels and touchpoints [5, 11–13]. Yet, particularly traditional BaM retailers that only have the stationary sales channel lack information regarding which e-service touchpoints they should provide to their customers [30]. It remains unclear, which e-service touchpoints potentially yield value for customers and would likely be used by them [31]. So far, there is a scarcity of empirical studies on omni-channel retailing [13], and only a limited amount of studies gives advice regarding digital technologies and e-service touchpoints to integrate into the physical servicescape [e.g., 3, 14, 22, 24].

This article suggests *mirroring*—complementing a traditional BaM retail channel with e-service touchpoints that have been adapted from proven designs in other channels. The lens of the dominant design theory provides us with a starting point for the investigation. The theory argues that a product category establishes a representative set of functions over time, which is then seen as standard [32]. The lens has previously been applied to technological milestones such as microprocessor designs, PC operating systems, and television systems [32, 33]. The dominance of a technology can be investigated on different levels of analysis. One level is the consideration of "technological artifacts as composed of subsystems that are linked together [...] through specific interfaces" [33, p.274]. In our context, these artifacts can comprise the whole omni-channel system or a particular digital channel such as an e-commerce solution. As e-commerce is a mature domain [3], it is fair to assume that leading e-commerce systems have set such a dominant design, which comprises a set of functions that represent the consolidated requirements of various types of users [32]. In the consequence, the e-service touchpoints offered by leading e-commerce systems should meet customers' requirements to a large extent and provide value for them. Hence, they also might be relevant for customers shopping in the physical BaM retail servicescape.

While the mirroring approach appears promising for BaM retailers, it remains unclear, if and to what extent customers are likely to use these mirrored touchpoints. Particularly, Suárez reminds us that "a dominant design is not always that design which has greatest technological sweetness" [32, p.417].

3 Mirroring E-service Touchpoints for Brick and Mortar Retail

3.1 Identification of Dominant E-service Touchpoints

We assessed leading e-commerce solutions to identify the e-service touchpoints commonly offered in e-commerce. This assessment covers both the propriety solutions of the three German e-commerce market leaders (*amazon.de, otto.de,* and *zalando.de*) [34] and the five leading commercial off-the-shelf e-commerce solutions in the German market (*Shopify, Magento, WooCommerce, XT:Commerce* and *Shopware*) [35]. As the latter solutions can be customized by the retailers to serve their needs, we selected concrete instances. To cater for potential variances resulting from national peculiarities, customizing, and different product categories offered, we sampled three major Western European e-commerce retailers for each solution, yielding a sample of 18 online shops.

We employed a qualitative web-content analysis approach [36], where two of the authors with backgrounds in retail and service science manually simulated typical customer journeys on the online shops. In an open coding process, we inductively derived a total of 35 different e-service touchpoints that guide, support or enable the customer during its customer journey. Since we focus on the dominant design, we only kept those touchpoints that were offered by at least four online shops. Further, logistics-related touchpoints, as well as ones that require an online shop (e.g., *Click & Collect* and *Click & Reserve*), have been eliminated, because they do not constitute an independent BaM e-service touchpoint. 19 unique e-service touchpoints remained.

Figure 1 shows the identified e-service touchpoints from e-commerce and maps them to their mirrored BaM e-service counterparts. Some e-service touchpoints were mirrored into more than one touchpoint, whereas other e-service touchpoints have been merged into a single mirrored touchpoint, yielding a total of 20 mirrored BaM e-service touchpoints. Based on the type of value proposition, we grouped the touchpoints into four distinct categories that occur along the customer journey, viz., search and navigation, product information, selection and checkout, and communication and support.

3.2 Mirrored E-Service Touchpoints

Search and Navigation Touchpoints are research-supporting shopping aids [37] that help customers in the pre-purchase phase of their journey [12] to find and locate relevant products and services from the retailer's offered portfolio. These touchpoints reduce the search time and potential information overload customers might experience.

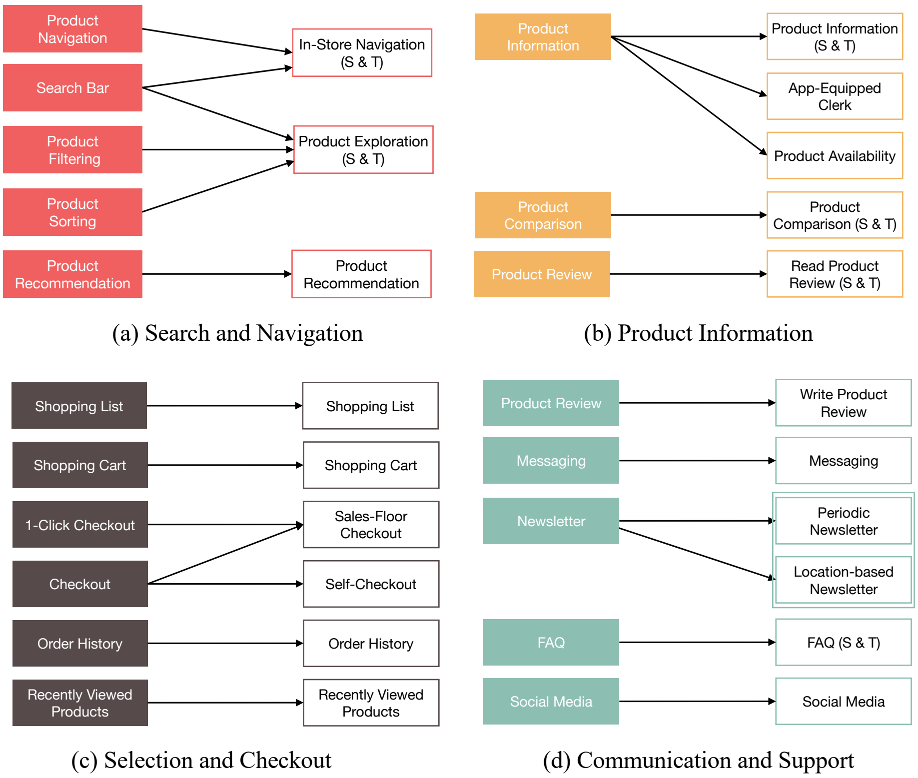


Figure 1. Mirroring of e-commerce touchpoints to BaM e-service touchpoints, grouped by touchpoint category (e-commerce touchpoint $\xrightarrow{\text{mirroring}}$ mirrored BaM e-service touchpoint)

In-Store Navigation Touchpoint: 83 % of the e-commerce solutions under consideration provide a global text search to directly track down desired products. This touchpoint is typically enhanced by auto-completion to assist with spelling and to provide first results before the search request is completed. Mirrored to the BaM context, this e-service touchpoint is particularly useful for large department stores and for customers unfamiliar with the store layout. It can support customers in locating a desired product and provides them with in-store directions. This touchpoint is feasible for both in-store terminals and smartphones. A smartphone-hosted app may use the tracking capabilities of the underlying device to provide real-time turn-by-turn navigation to a product.

Product Exploration Touchpoint: Due to their typically extensive product portfolios, online shops provide various means to browse and explore the portfolio. Sorting (e.g., lowest price or best reviews first) and filtering (e.g., product categories, brands) capabilities are provided to narrow down the product portfolio. Even though BaM stores fall short in terms of portfolio sizes when compared to online retailers [38], most retailers still offer a wide range of products. While e-commerce search touchpoints give direct product access, the portfolio might be harder to perceive in-store due to product presentation. In effect, potential BaM customers can experience issues in their orientation and selection process. Mirroring existing product exploration touchpoints to the BaM environment is facilitated by accessing the retailer's product data through terminals and smartphones, allowing the customer to apply the known filtering and sorting operations.

Product Recommendation Touchpoint: More than two-thirds of the considered e-commerce solutions feature a product recommendation engine. Based on behavioral customer data such as order history and recently viewed products, the e-service recommends products similar to previously considered items (content filtering) or ones that customers with similar taste have bought (collaborative filtering) [39]. By now, recommended products make up to 35 % of e-commerce purchases [40]. A mirrored product recommendation service takes into account behavioral customer data such as order history, shopping lists, or even their current in-store location. A smartphone app collects this information and, in turn, makes product recommendations.

Product Information Touchpoints assist the customer with retrieving of detailed information about considered products, and aid in the product selection process. As such, they primarily take place in the pre-purchase phase of the customer journey [12].

Extended Product Information Touchpoint: Urged by the inability of customers to physically experience products online, retailers responded by providing extensive textual and multi-media product presentations, which has led to online retailers giving more detailed information than their BaM counterparts. Customers can access a mirrored product information touchpoint by scanning product tags with their smartphones to receive detailed multi-media product information. Similarly, customers can carry products to an in-store terminal, scan their tags, and retrieve the information.

App-equipped Clerk Touchpoint: Some customers prefer personal service over e-service to get product information. Typically, store associates must have basic knowledge of all products in their responsible department. In case of detail questions, they might not be able to provide a correct answer. By providing store associates with a product information touchpoint (e.g., via a smart device), they can quickly look up the requested information for the customer.

Product Availability Touchpoint: Two in three analyzed online shops inform on product availability, i.e., whether a selected item is in stock, how many pieces are left, and send notifications when an out-of-stock product becomes available again. In a BaM store, product availability is visually apparent. However, even if a product is not on display, it might be available in the back room or a nearby warehouse. Through the introduction of a product availability touchpoint, which can be accessed by scanning a product's shelf label or manually searching a product, customers can access backstage warehouse availability information, trigger a refill or get a refill notification. As such, frustration from out-of-stock situations can be mitigated.

Product Comparison Touchpoint: As seen in the extended product information touchpoint, many online stores allow to compare the information linked to different products (e.g., size, price, and technical features) in a tabular fashion to improve the customer's selection process. Hence, a mirrored e-service touchpoint may allow scanning the tags of multiple products with a smartphone or in-store terminal app in order to obtain structured information for an efficient comparison.

Read Product Review Touchpoint: Product reviews, typically written by other customers, are one of the major information sources for online shoppers in their decision process [41]. The touchpoint can be mirrored as an extension to the product information service. In effect, the collective experience of the community in the form of ratings and reviews can be shown next to the retailer-provided product information.

Selection and Checkout Touchpoints support customers by helping them to plan future purchases, retaining control of the current purchase process, facilitating the checkout process, and keeping track of the past purchases [12].

Shopping List Touchpoint: E-commerce solutions frequently allow registered customers to create shopping lists or wish lists to save product references for later consideration. Interestingly, this touchpoint does not have its origins in the digital realm but was previously mirrored from traditional paper-based lists from the offline world. Now, the e-service implementation can be mirrored back to the physical realm. Customers can use a smartphone app to add selected products to a shopping list. Besides remembering relevant products, shopping lists allow for easy access to recurring purchases and to prepare store visits in advance.

Shopping Cart Touchpoint: Similar to shopping lists, shopping carts first had a physical representation before getting implemented in e-commerce. Digital shopping carts (or shopping bags) contain all selected products, display the subtotal and commonly allow to search and filter their contents. Mirrored back as an e-service for BaM, customers scan selected products before placing them into the physical shopping cart. Using their smartphones, customers can keep track of their projected spending and compare their selection to previously created shopping lists.

Sales-Floor Checkout Touchpoint: In e-commerce, customers do not have to wait in a queue or depend on a cashier to complete the checkout process, which renders it very efficient. One way to mirror a more efficient checkout to the physical realm lies in equipping store associates with smart devices to scan products and check out the customer on the sales floor. Hereby, the queue at the sales counter can be skipped.

Self-Checkout Touchpoint: Another way to digitally increase checkout efficiency is self-checkout. While these systems originally were introduced as a means for retailers to reduce labor cost [42], they can also facilitate efficiency gains for customers (e.g., skipping queues). In conjunction with the shopping cart touchpoint, a mobile self-checkout becomes feasible, where customers would only have to complete their purchase with electronic payment. This touchpoint satisfies shoppers' demand for "sofortness", similar to what they are used to from e-commerce.

Order History Touchpoint: In online shops, registered users usually can access prior orders to find out what has been bought. Mirrored to a smartphone app in the physical realm, a digital order history can help to keep an overview of what has recently been bought, and also to support recurring purchases such as groceries. The touchpoint also supports keeping track of the spending over time.

Recently Viewed Products Touchpoint: During online shopping, customers often interact with dozens if not hundreds of different products, which makes it hard to keep track of the items recently viewed. Respective touchpoints support customers by caching the accessed items. While BaM store portfolios are not as extensive as their online counterparts, they might still be sufficiently large to make it difficult for customers to keep track of all viewed products and retrieve the most desirable one (e.g., make a choice from different tried-on clothes). Again, customers can use their smartphone to scan the tags of considered products.

Communication and Support Touchpoints are used throughout the whole customer journey [12] to facilitate marketing, customer engagement, and customer care.

Write Product Review Touchpoint: Besides customers digitally consuming reviews, providing ratings and reviews themselves can also be mirrored by means of a smartphone app to the BaM context. By providing their opinion, customers support other shoppers and may also help retailers to adjust their product portfolio w.r.t. the customer reviews.

Messaging Touchpoint: Even though online retailers do not have the traditional shop-floor personnel, many of them provide customer service by means of a (video-)chat interface. Mirrored to a BaM store, an instant messaging touchpoint can ease interactions with store associates, especially when no associate is nearby.

Newsletter Touchpoint: Newsletters typically inform about a retailer's special offers or events. In contrast to conventional mass mailing such as catalogs, digital newsletters can be personalized. Personalized digital newsletters in a BaM environment can be viewed by the customer through a smartphone app. Further, a location-based newsletter touchpoint can exploit the smartphone's technological capabilities to enable location-based proactive newsletters that are sent out when a customer is passing by the store.

FAQ Touchpoint: Beyond product-related information, online retailers typically provide a list of answers to so-called "Frequently Asked Questions" (FAQ), which deal with topics such as warranty handling, delivery times, return policies or payment methods. Since FAQs are not customer-specific, this touchpoint can be mirrored both via smartphone and in-store terminal. An FAQ touchpoint helps to clarify time-consuming questions in advance that might otherwise be asked during consultations with store associated or at the checkout counter.

Social Media Touchpoint: At its beginnings, online shopping was a mere functional activity without a prevalent social component. With the rise of social media, the concept of "social commerce" emerged [43], where customers share their shopping experiences with others. While customers in BaM can directly interact with friends and other shoppers in-store, access to social media would enable interaction with a much larger group of acquaintances and strangers online. A retailer-provided social media touchpoint can, for example, facilitate discussions between customers or enable customers to showcase newly bought items.

4 Assessment of Mirrored E-Service Touchpoints

4.1 Survey Research Approach

Quantitative survey research [44] was employed to identify the shoppers' likeliness of use regarding the mirrored e-service touchpoints. We conducted a hypothetical thought experiment where participants were asked to imagine being in a fictitious "smart store", instead of presenting and surveying particular instances of the envisioned touchpoints. In an online survey, the participants were first introduced to the study context and the four categories of touchpoints that the fictional smart store will offer to complement traditional store operations (see Section 3). Afterward, each e-service was briefly introduced in one paragraph. Given the nature of the study where touchpoints were presented on a rather abstract level, we did not apply measures such as the Technology Acceptance Model [45], which are focused on well-specified particular instances of technology. Instead, for each touchpoint, we surveyed its likeliness of use by means of a single item five-point Likert-scale (e.g., "*how likely would it be for you to use such a product availability service?*"). When applicable, the question was asked twice—both for accessing the touchpoint in-store through a smartphone and through an in-store terminal. Finally, participants had to provide demographic information. In light of the overall relevance of BaM retail in society, we did not limit the survey to a particular audience. Instead, we recruited a diverse sample of more than 300 participants of at least 18 years of age from the Western world using *prolific.ac*. The recruiting platform aims at providing researchers with representative samples of users regarding age, gender, and educational level. An attention check question (adapted from [46]) was used to filter inattentive participants, which yielded 250 valid responses. The mean age of the respondents was 32.69 years (median 31 years, SD 9.96 years) and 53.60 % were female. Participants countries of origin were the United Kingdom (145), the United States (52), Canada (12), Portugal (9), the Netherlands (4) and sixteen other European countries (28). The average completion time for the survey was 9 minutes and 37 seconds. Participants were rewarded £1.15.

4.2 Survey Results

Figure 2 provides the distribution of the likelihood of use of the mirrored e-service touchpoints. The touchpoints are grouped by the four categories, and within each category, the e-service touchpoints are sorted in descending order by their average rating. The global average rating is 3.58 points. The social media service ranked worst with an average of 2.23 points, while the self-checkout service ranked best with an average of 4.35 points. An inter-group comparison revealed that respondents are most likely to use selection and checkout touchpoints (Ø 3.92 points). Product information touchpoints (Ø 3.79 points) and search and navigation touchpoints (Ø 3.73 points) ranked similarly, whereas respondents are much less likely to use communication and support touchpoints (Ø 2.94 points). The social media ($\sigma^2 = 1.71$; $\sigma = 1.31$), periodical newsletter ($\sigma^2 = 1.64$; $\sigma = 1.28$), location-based newsletter ($\sigma^2 = 1.60$; $\sigma = 1.27$), and messaging touchpoints ($\sigma^2 = 1.48$; $\sigma = 1.22$) were also the most controversial ones.

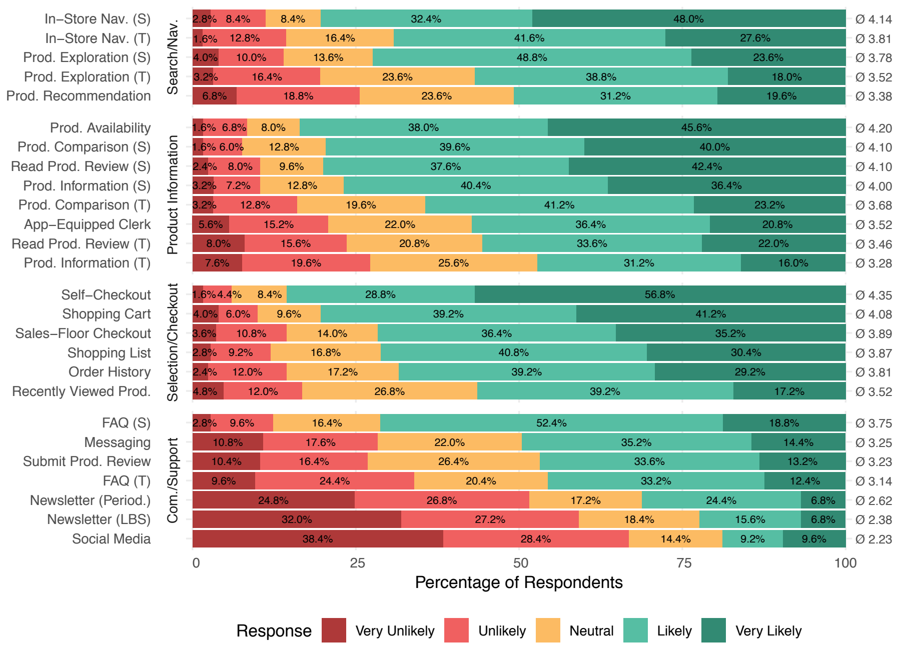


Figure 2. Distribution of likelihood of use per e-service touchpoint (grouped by touchpoint categories; within groups ordered descending by average score)

Pairwise Spearman's rank coefficients ρ between each e-service and the respondents' age indicated no significant relationships for most touchpoints, except for product comparison via in-store terminal ($\rho = .48$; $p < .001$), reading product reviews via in-store terminal ($\rho = .45$; $p < .01$), and accessing FAQs on an in-store terminal ($\rho = .39$; $p < .01$), which all show a positive relationship to increased age. On the contrary, the social media service ($\rho = -.43$; $p < .01$) has a negative relationship to increased age. Regarding relationships between the respondents' gender and their answers, Kendall's τ_b did not indicate significant relationships, except for a weak positive relationship between female respondents and the messaging service ($\tau_b = .16$; $p < .01$). However, women on average voted .08 points higher than men. Although not statistically significant, women, in particular, had a higher likeliness to use a smartphone (.31 points) or an in-store terminal (.38 points) to read product reviews compared to male respondents.

Six e-service touchpoints were surveyed regarding the two service interfaces smartphone (denoted (S) in Figure 2) and in-store terminal (denoted (T) in Figure 2). Results show that respondents prefer smartphones over terminals by an average of .50 points, and every e-service was ranked higher when accessed via a smartphone app. Differences were lowest for the product exploration service (.26 points) and highest for the extended product information service (.71 points). No unfamiliarity effects regarding mobile devices and e-commerce were evident, which could have influenced participants' responses. 99.20 % of respondents report owning a smartphone and 182 participants (72.80 %) report using their smartphone or tablet in-store. All but one respondent (99.60 %) have purchased goods and services online at least once. Nine participants (3.60 %) use e-commerce on a daily basis; 105 participants (42.00 %) report weekly e-commerce activities; whereas every second participant roughly shops online once a month. Ten participants use e-commerce as infrequently as once a year.

5 Discussion, Limitations, and Future Research

We mirrored e-service touchpoints for their application in BaM, and assessed shoppers' likeliness to use them to complement traditional service. Our findings offer several implications for academia and practice.

First, an overall positive response towards complementary e-service touchpoints provides support for channel complementary theory in the BaM retail domain [29] and supports the assumption that customers have a positive sentiment towards e-service in BaM retail. The results show that customers are likely to use e-service touchpoints in the physical retail servicescape to complement and improve their customer journey. In particular, selection and checkout touchpoints have a very positive response. It is fair to assume that most customers are used to the well-rated self-checkout touchpoint [47] and a mere-exposure effect might have occurred, i.e., respondents rated the touchpoint well just because of familiarity [48]. Nevertheless, most well-rated touchpoints comprise research-supporting and solution-oriented shopping aids [37], which increase customers' shopping effectiveness (find the right products) [22] and efficiency (fast service with high quality) [31, 49].

Next, the customers' age has a positive impact on the likeliness to use in-store terminals instead of smartphone apps. Also, it has a negative impact on the use of social media. These findings support prior studies on the use and adoption of self-service technology [42, 47, 49]. According to these studies, technology anxiety and effort expectancy have been found to determine the intention to use self-service technology. In-store terminals closely resemble other well-established self-service technologies such as ATMs or ticket machines that are common for a few decades. Consequently, also older customers are familiar with these types of technology and might be more confident in using them [47]. The proliferation of smartphones, on the contrary, has just started in the last decade, which implies comparably less experience with this technology [42] and potentially higher technology anxiety [49]. Nevertheless, customers overall prefer smartphone apps to access e-service touchpoints. While location-independence, the degree of customer-retailer connectivity, and perceived control over the shopping process are known influences on the use of smartphone apps in retail [21], future research is needed to clarify the determinants of in-store smartphone app usage. Lastly and in line with Yoo and Gretzel [50], female customers are more motivated to provide product reviews (e.g., to help other customers).

As with any research, our work comes with some limitations. Inherent to the survey research method are potential negative side effects or biases. Considering the rather "demanding" responses from the survey, some respondents might suffer from a "good-subject effect" where participants try to guess the purpose of a survey to give pleasing answers [51]. Additionally, we did not survey participants' prior experiences with the non-mirrored touchpoints, which will likely have influenced individual responses. This leaves room for further research on the influence of prior experiences in one channel on other complementary channels. As regards the unit of analysis of our study, we assessed the likeliness of use regarding the mirrored e-service touchpoints, without providing concrete touchpoint implementations and contextual factors. The relevance of e-service touchpoints is likely to differ, among others, based on the categories of goods sold and the size of the store. For example, an in-store navigation service might be more useful in a large fashion store that sells a plethora of different articles, than in a small delicatessen store where customers can find products easily without digital support. Also, the relevance of product information touchpoints might also be related to the complexity of the product, i.e., customers might require less information on a broom than on a robot vacuum cleaner. Nevertheless, our assessment provides a foundation and justification for BaM retailers to conduct individual, detailed analyses of e-service touchpoints that are feasible and economically viable for them.

Omni-channelling and e-service touchpoints in BaM are still emerging and developing at a fast pace [14, 24], leaving ample room for further research. First, retailers require instruments for decision support regarding the introduction of e-service touchpoints, as there is a variety of opportunities that can be considered by BaM retailers. In particular, SME retailers with limited financial resources require support, as high the upfront cost is known as a strong inhibitor for the adoption of digital channels [7]. This raises the question of whether the introduction of novel e-service touchpoints is feasible for SME BaM retailers. Further empirical work is required to investigate retailer-sided adoption and customer acceptance of e-service touchpoints in BaM retail with special consideration on the categories of goods sold, the size of the store, and the retailer's competitive strategy (e.g., individualization vs. cost optimization) [22]. Such research could result in a decision calculus that embeds impact, adoption, and success factors. Second, besides in-store terminals and smartphone apps, there are other digital technologies that enable the implementation of e-service touchpoints in the physical servicescape [14, 22, 24]. Further research could compare and contrast these partly complementary technologies w.r.t. the e-service touchpoints provided and effects on the customers observed. Third, in the long run, a set of common e-service touchpoints in the physical servicescape might eventually prevail. Using the multiple case study method, research can then identify a dominant design [32] of BaM e-service touchpoints to further support touchpoint management in BaM retail. However, even with such a dominant design, ongoing research is needed due to the continuous nature of the digital transformation [1]. As illustrated in our discussion of the different e-services in Section 3, e-commerce and BaM retail mutually stimulate each other in terms of digital innovation. First, e-commerce tried to introduce touchpoints similar to those known from BaM (e.g. social components). Now, e-service touchpoints get introduced to the physical servicescape. This interplay provides an interesting area for ongoing investigations.

6 Conclusion

Given that many traditional BaM retailers struggle to keep up with the digital transformation and the induced competitive disadvantage, this paper offers two primary contributions to research and practice: First, we applied dominant design theory to identify major e-service touchpoints that are commonly implemented in leading e-commerce software solutions, grouped them according to their value proposition, and mirrored them to complement the physical servicescape of BaM. Second, we assessed these touchpoints by means of an online survey. In support of channel complimentary theory, results showed that customers are likely to use e-service touchpoints in-store. Customers preferred touchpoints that aid in product search, information, and selection as well as facilitate an efficient customer journey. Communication, social and support services, on the contrary, were surprisingly undesirable. In line with related studies, in-store terminals have been identified as the preferred choice of the older generation, while smartphones are the overall favorable method to access e-service. Future research can build upon these insights to design and implement pilots for the most promising e-service touchpoints together with retail organizations. Subsequent empirical studies can then shed light on the advantages provided by, and customer acceptance and use of these mirrored e-service touchpoints.

Acknowledgments

This paper was developed in the research project *smartmarket*[2], which is funded by the German Federal Ministry of Education and Research (BMBF), promotion sign 02K15A074. The authors thank the Project Management Agency Karlsruhe (PTKA).

References

1. Matt, C., Hess, T., Benlian, A.: Digital Transformation Strategies. Bus. Inf. Syst. Eng. 57(5), 339–343 (2015)
2. Gray, J., Rumpe, B.: Models for the digital transformation. Softw. Syst. Model. 16(2), 307–308 (2017)
3. Kumar, V., Anand, A., Song, H.: Future of Retailer Profitability: An Organizing Framework. J. Retail. 93(1), 96–119 (2017)
4. Hagberg, J., Sundstrom, M., Egels-Zanden, N.: The Digitalization of Retailing: an Exploratory Framework. Int. J. Retail Distrib. Manag. 44(6), 336–368 (2016)
5. Verhoef, P.C., Kannan, P.K., Inman, J.J.: From Multi-Channel Retailing to Omni-Channel Retailing. Introduction to the Special Issue on Multi-Channel Retailing. J. Retail. 91(2), 174–181 (2015)
6. Bollweg, L., Lackes, R., Siepermann, M.,Weber, P.: The Role of E-Intermediaries in Local Retail Hyperlink Networks: A Hyperlink Network Analysis. In: Proc. MKWI 2018. pp. 514–525. Lüneburg, Germany (2018)
7. Betzing, J.H., Beverungen, D., Becker, J.: Design Principles for Co-Creating Digital Customer Experience in High Street Retail. In: Proc. MKWI 2018. pp. 2083–2094. Lüneburg, Germany (2018)

8. Statista: Handel: Offline vs. Online: Entwicklung, Vergleiche, Konsumenten (2017), https://de.statista.com/statistik/studie/id/43363/dokument/, last accessed 2018-11-19
9. Heinemann, G.: Die Neuerfindung des stationären Einzelhandels: Kundenzentralität und ultimative Usability für Stadt und Handel der Zukunft. Springer Gabler, Wiesbaden, Germany (2017)
10. Eichholz-Klein, S., Preißner, M., Lerch, C., Brylla, T.: Stadt, Land, Handel 2020. Tech. rep., IFH Köln, Cologne, Germany (2015)
11. Saghiri, S., Wilding, R., Mena, C., Bourlakis, M.: Toward a three-dimensional framework for omni-channel. J. Bus. Res. 77, 53–67 (2017)
12. Lemon, K.N., Verhoef, P.C.: Understanding Customer Experience and the Customer Journey. J. Mark. 80(6), 69–96 (2016)
13. von Briel, F.: The future of omnichannel retail: A four-stage Delphi study. Technol. Forecast. Soc. Change 132, 217–229 (2018)
14. Inman, J.J, Nikolova, H.: Shopper-Facing Retail Technology: A Retailer Adoption Decision Framework Incorporating Shopper Attitudes and Privacy Concerns. J. Retail. 93(1), 7–28 (2017)
15. Cahn, J.: Don’t Let Your Industry, Job Or Company Get Amazon-ed (2018), https://www.forbes.com/sites/jamescahn/2018/04/24/dont-let-your-industry-job-or-company-get-amazon-ed/, last accessed 2018-11-19
16. Lu, Q.S., Pattnaik, C., Xiao, J., Voola, R.: Cross-national variation in consumers’ retail channel selection in a multichannel environment: Evidence from Asia-Pacific countries. J. Bus. Res. 86, 321–332 (2018)
17. Vargo, S.L., Lusch, R.F.: Why ’service’? J. Acad. Mark. Sci. 36(1), 25–38 (2008)
18. Rowley, J.: An analysis of the e-service literature: towards a research agenda. Internet Res. 16(3), 339–359 (2006)
19. Heuchert, M., Barann, B., Cordes, A.K., Becker, J.: An IS Perspective on Omni-Channel Management along the Customer Journey: Development of an Entity-Relationship-Model and a Linkage Concept. In: Proc. MKWI 2018. pp. 435–446. Lüneburg, Germany (2018)
20. Halvorsrud, R., Kvale, K., Følstad, A.: Improving service quality through customer journey analysis. J. Serv. Theory Pract. 26(6), 840–867 (2016)
21. Faulds, D.J., Mangold, W.G., Raju, P., Valsalan, S.: The mobile shopping revolution: Redefining the consumer decision process. Bus. Horiz. 61(2), 323–338 (2018)
22. Betzing, J.H., Hoang, A.Q.M., Becker, J.: In-store Technologies in the Retail Servicescape. In: Proc. MKWI 2018. pp. 1671–1682. Lüneburg, Germany (2018)
23. Voorhees, C.M., Fombelle, P.W., Gregoire, Y., Bone, S., Gustafsson, A., Sousa, R., Walkowiak, T.: Service encounters, experiences and the customer journey: Defining the field and a call to expand our lens. J. Bus. Res. 79, 269–280 (2017)
24. Willems, K., Smolders, A., Brengman, M., Luyten, K., Schöning, J.: The path-topurchase is paved with digital opportunities: An inventory of shopper-oriented retail technologies. Technol. Forecast. Soc. Change 124, 228–242 (2017)
25. Straker, K., Wrigley, C., Rosemann, M.: Typologies and touchpoints: designing multi-channel digital strategies. J. Res. Interact. Mark. 9(2), 110–128 (2015)
26. Nüesch, R., Alt, R., Puschmann, T.: Hybrid Customer Interaction. Bus. Inf. Syst. Eng. 57(1), 73–78 (2015)
27. Hosseini, S., Röglinger, M., Schmied, F.: Omni-Channel Retail Capabilities: An Information Systems Perspective. In: Proc. ICIS 2017. Seoul, South Korea (2017)
28. Herhausen, D., Binder, J., Schoegel, M., Herrmann, A.: Integrating Bricks with Clicks: Retailer-Level and Channel-Level Outcomes of Online-Offline Channel Integration. J. Retail. 91(2), 309–325 (2015)

29. Dutta-Bergman, M.J.: Interpersonal communication after 9/11 via telephone and internet: A theory of channel complementarity. New Media Soc. 6(5), 659–673 (2004)
30. Wagner, G., Schramm-Klein, H., Steinmann, S.: Effects of cross-channel synergies and complementarity in a multichannel e-commerce system. Int. Rev. Retail. Distrib. Consum. Res. 23(5), 571–581 (2013)
31. Blázquez, M.: Fashion Shopping in Multichannel Retail: The Role of Technology in Enhancing the Customer Experience. Int. J. Electron. Commer. 18(4), 97 – 116 (2014)
32. Suárez, F.F., Utterback, J.M.: Dominant designs and the survival of firms. Strateg. Manag. J. 16(6), 415–430 (1995)
33. Suárez, F.F.: Battles for technological dominance: an integrative framework. Res. Policy 33(2), 271–286 (2004)
34. Statista: B2C-E-Commerce: Ranking der Top 100 größten Online-Shops nach Umsatz in Deutschland im Jahr 2016 (2016), https://de.statista.com/statistik/daten/studie/170530/, last accessed 2018-11-19
35. Datanyze: E-Commerce Platforms Market Share Germany (2018), https://www.datanyze.com/market-share/e-commerce-platforms/Germany, last accessed 2018-11-19
36. Mayring, P.: Qualitative Content Analysis. Theoretical Foundation, Basic Procedures and Software Solution. Beltz, Klagenfurt, Austria (2014)
37. Angela Chang, C., Kukar-Kinney, M.: Effects of shopping aid usage on consumer purchase decision and decision satisfaction. Asia Pacific J. Mark. Logist. 23(5), 745-754 (2011)
38. Brynjolfsson, E., Hu, Y.J., Simester, D.: Goodbye Pareto Principle, Hello Long Tail. Manage. Sci. 57(8), 1373–1386 (2011)
39. Balabanovi´c, M., Shoham, Y.: Content-Based, Collaborative Recommendation. Commun. ACM 40(3), 66–72 (1997)
40. MacKenzie, I., Meyer, C., Noble, S.: How retailers can keep up with consumers. Tech. rep., McKinsey & Company (2013)
41. Floyd, K., Freling, R., Alhoqail, S., Cho, H.Y., Freling, T.: How Online Product Reviews Affect Retail Sales: A Meta-analysis. J. Retail. 90(2), 217–232 (2014)
42. Dean, D.H.: Shopper age and the use of self-service technologies. Manag. Serv. Qual. 18(3), 225–238 (2008)
43. Hajli, N.: A study of the impact of social media on consumers. Int. J. Mark. Res. 56(3), 387 (2014)
44. Recker, J.: Scientific Research in Information Systems: A Beginner's Guide. Springer, Berlin, Heidelberg (2012)
45. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q. 13(3), 319 (1989)
46. Oppenheimer, D., Meyvis, T., Davidenko, N.: Instructional manipulation checks: Detecting satisficing to increase statistical power. J. Exp. Soc. Psychol. 45(4), 867–872 (2009)
47. Wang, Y.S., Shih, Y.W.: Why do people use information kiosks? A validation of the Unified Theory of Acceptance and Use of Technology. Gov. Inf. Q. 26(1), 158–165 (2009)
48. Zajonc, R.B.: Attitudinal Effects of Mere Exposure. J. Pers. Soc. Psychol. 9(2), 1–27 (1968)
49. Meuter, M.L., Ostrom, A.L., Bitner, M.J., Roundtree, R.: The influence of technology anxiety on consumer use and experiences with self-service technologies. J. Bus. Res. 56(11), 899–906 (2003)
50. Yoo, K.H., Gretzel, U.: What Motivates Consumers to Write Online Travel Reviews? Inf. Technol. Tour. 10(4), 283–295 (2008)
51. Nichols, A.L., Maner, J.K.: The Good-Subject Effect: Investigating Participant Demand Characteristics. J. Gen. Psychol. 135(2), 151–166 (2008)

Taxonomy of Digital Platforms: A Platform Architecture Perspective

Michael Blaschke[1], Kazem Haki[1], Stephan Aier[1], and Robert Winter[1]

[1] University of St.Gallen, Institute of Information Management, St.Gallen, Switzerland
{michael.blaschke,kazem.haki,stephan.aier,robert.winter}
@unisg.ch

Abstract. Digital platforms—technical core artefacts augmented by peripheral third-party derivatives—afford organizations to integrate resources in networked business ecosystems. Although digital platforms widely differ in their configurations, digital platforms' *dimensions* and *characteristics* to disentangle different digital platform configurations are under-researched. To bridge this void, we employ Nickerson et al.'s method for taxonomy development to systematically derive a taxonomy of digital platforms. Specifically, we embrace a platform architecture perspective to capture the configuration of digital platform's components. The resultant taxonomy facilitates a more pronounced understanding and grouping of digital platforms as configurations of certain *dimensions* and *characteristics*. Our findings suggest that digital platforms exhibit characteristics on at least four dimensions—namely, *infrastructure*, *core*, *ecosystem*, and *service* dimensions. Second, through instantiating the taxonomy, we find that digital platforms that exhibit *similar* characteristics share identical architectural profiles and, therefore, belong to one of three digital platform archetypes—namely, *orchestration*, *amalgamation*, and *innovation* platforms.

Keywords: Digital Platforms, Taxonomy, Platform Architecture, Platform Ecosystems, Archetypes.

1 Introduction

This study investigates *digital platforms*—sets of stable technical core artefacts augmented by peripheral third-party derivatives, and associated organizational arrangements [1]. A digital platform facilitates the integration of resources in business ecosystems and becomes increasingly valuable when more third parties join the platform and add their complementary derivatives [2]. Omnipresent in today's industries, digital platforms differ in their configurations [2, 3]—as exemplified by social media (e.g., Facebook and LinkedIn), mobile operating system (e.g., Android and iOS), payment (e.g., PayPal and Apple Pay), and peer-to-peer (e.g., Uber and Airbnb) platforms.

Beyond the diversity of digital platforms in practice, our review of digital platform literature exposes a wide variety of digital platform conceptualizations [1, 4]. We are specifically concerned that IS and management discourses on digital platforms [4] do

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

not consider the specific characteristics of digitality [5]. Conversely, they treat all technological platforms as a homogeneous group in which classifications are merely based on organizational arrangements [1]. For example, for *digital* platforms, openness does not merely relate to organizational arrangements such as entrance and exit rules, but also to openness of technologies such as software development kits [1].

The abovementioned diversity in digital platforms' instances and conceptualizations calls for a digital platform taxonomy to disentangle different digital platform configurations [2]. Taxonomies play a vital role in research and practice because the classification of objects helps researchers and practitioners understand and analyze complex domains [6]. For digital platforms, a taxonomy would organize digital platforms' divers instances and conceptualizations into a coherent organizing structure. To this end, we first extract digital platforms' *dimensions* and *characteristics* from existing digital platform instances and studies. Relying on such dimensions and characteristics, we then develop a digital platform taxonomy and eventually instantiate the resultant taxonomy to derive digital platform archetypes. This research therefore seeks to answer the following research question: *Which dimensions and characteristics distinguish digital platforms through their architectural configuration?*

To answer the research question, we first follow Reuver et al.'s [1] recommendation to provide clear definitions for key concepts in the digital platform context. Subsequently, we follow Nickerson et al.'s step-by-step and well-structured method for taxonomy development [6]. In this process, we code digital platform articles to identify a sample of 34 digital platform instances. The resultant taxonomy postulates digital platforms' dimensions and characteristics. We instantiate this taxonomy with the 34 digital platform instances to derive digital platform archetypes that capture archetypical configurations of digital platform profiles with similar characteristics.

Thereunto, we promote the use of *platform architecture* as a focused perspective to effectively capture the configuration of a given digital platform's components. Platform architecture here refers to the fundamental organization of a digital platform, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution [4, 7]. We opt for this perspective as it conceptualizes digital platforms as layered modular architectures that uniquely differ in their components' configurations. Relying on the platform architecture perspective, we supplement prior research with a taxonomy and archetypes of digital platforms both of which rest on digital platforms' architectural dimensions and characteristics.

2 Research Background

Since this study aims to develop a taxonomy of digital platforms, we first provide an overview of digital platform research. We then review the presence of taxonomies in IS (in general) and digital platform research (in particular) to motivate and position our study. Eventually, we introduce digital platform architecture as this study's specific perspective. In *briefly* sketching these streams to examine their underlying logic, our citations to these vast streams are merely illustrative; a thorough review of each would be a substantial and worthwhile project in its own right.

2.1 Digital Platforms

Originally viewed as facilitator of *bilateral* innovation activities (late 1990s) [e.g., 8], the platform concept increasingly captured networked, multi-lateral innovation activities (mid-2000s) [e.g., 9]. IS research then studied platforms as a *central* form of organizing technological innovation (2010s) [e.g., 10]. Today, the term *platform* is omnipresent in both IS and management research [1-4, 11], such as the *Information Systems Research* (ISR) special issues on *Platforms and Infrastructures in the Digital Age* [5]. Thomas et al. [4] organize platform research on a continuum from *firm-internal* to *firm-external* platforms. As digital platforms represent "layered modular technology architectures in business networks" [3, p. 186], they lie on the *firm-external* end of platform research that spotlights such business *networks* (e.g., *Android*'s mobile ecosystem). Within these networks, digital platforms mediate actor-to-actor interactions [2] and leverage innovation [12]. We thus view digital platforms as a *socio-technical* phenomenon rather than purely *technical* artefacts as they encompass both a technical core as well as business networks mediated by a technical core [1]. Table 1 synthesizes the key concepts that represent our understanding of digital platforms.

Table 1. Key Concepts in the Digital Platform Context

Concept	*Definition*
Platform Owner	Natural or legal entity that designs, implements, and maintains the digital platform [13]
Third Party	Natural or legal entity that augments the technical core with complementary derivatives (e.g., software extensions, services, and sales channels) [14]
End User	Natural or legal entity that uses the resources available on the digital platform [2]
Digital Ecosystem	Complex network of platform-mediated actor-to-actor interactions, turning increasingly accessible to end users through third parties' platform derivatives [13].
Service	Specialized competences (knowledge and skills) exchanged among different actors in the digital ecosystem through deeds, processes, and performances [15]
Technical Core	Extensible codebase serving as a building block upon which third parties devise platform-augmenting derivatives [16]
Digital Infrastructure	The computing and network resources that allow distributed actors to facilitate their resource exchange across spatial, temporal, and organizational boundaries [5]

2.2 The Role of Taxonomies in IS and Digital Platform Research

Intuitively, taxonomies[1] serve as sorting schemes to systematically organize objects in a domain of interest (e.g., digital platforms), a fundamental problem in many research disciplines [17, 18]. Technically, Nickerson et al. define a taxonomy T as a set of n dimensions, with each dimension consisting of at least two mutually exclusive and collectively exhaustive characteristics such that each object under consideration instanti-

[1] Prior literature often uses the different terms classification, framework, typology, and taxonomy equivalently [5]. As we employ Nickerson et al.'s method for *taxonomy* development [5], and as *taxonomy* is also the most common term across research disciplines, we opt for common recognition and consistency and use taxonomy *exclusively*.

ates one and only one characteristic for each dimension [6, p. 440]. The role of taxonomies—organizing IS domains through classifying objects of interest within these domains—is well recognized in the IS literature. Glass and Vessey [19] note that taxonomies provide an organizing structure to the IS body of knowledge. Fiedler et al. emphasize that taxonomies have been important in research "since Aristotelian applications over 2000 years ago" [20, pp. 11-12]. Similarly, Sabherwal and King argue that "taxonomies also help us understand divergence in previous research findings" [21, p. 180].

In the specific domain of digital platforms, prior digital platform research calls for using taxonomies for distinguishing digital platforms to ultimately specify different digital platform configurations [1, 3, 5]. However, only few theoretical accounts postulate *fragmented* dimensions and characteristics of digital platforms. For instance, while Kazan et al. [3, p. 187] conceptualize "two strategic architectural dimensions" of digital platforms—that is, (1) *core platform* and (2) *infrastructure* dimensions—, their research objective is not to classify digital platforms. Similarly, Williams et al. [22] focus on digital platforms' *digital service* dimension in deriving a taxonomy for platform-mediated digital services. Karhu et al. [23] promote *platform openness* as a dimension of platform architecture—differentiating *access openness* and *resource openness* (characteristics). However, their phenomenon of interest is platform forking in which a hostile firm (i.e., a forker) exploits a digital platform's shared resources, core and complements, to create a competing platform business. Overall, as there are fragmented discussions on classifying digital platforms, we reconcile a set of digital platform articles in the organizing structure of dimensions and characteristics to systematically derive a taxonomy of digital platforms for a specified use and purpose—that is, distinguishing digital platforms based on their architectural configuration.

2.3 Digital Platform Architecture

As the use of proper research perspectives guides IS scholars in both theory building and theory testing [24], in this study we promote the use of platform architecture as a purposeful research perspective to study configurations of digital platforms' components [7]. The targeted taxonomy's purpose is to distinguish digital platforms based on common *characteristics* within architectural *dimensions*. Our perspective effectively serves this purpose as follows. First, through viewing digital platforms as "layered modular technology architectures in business networks" [3, p. 186], this perspective accounts for the socio-technical and complex nature of digital platforms [1]. Second, its conception of digital platforms as layered modular technology architecture allows us to derive standalone but differentiating digital platform dimensions. Third, the platform architecture perspective describes a digital platform's architectural configuration to reflect the unique combination of a digital platform's components. Ultimately, this perspective facilitates the identification of digital platform archetypes as digital platforms exhibiting *similar* architectural configurations belong to the same archetype. Beyond these reasons, prior research also motivates the significance of using a platform architecture perspective for distinguishing digital platforms [1, 3, 5, 7].

3 Research Method

In this section, we outline the applied steps in our taxonomy development. We then instantiate the resultant taxonomy to derive digital platform archetypes.

Digital Platform Taxonomy. We adopt Nickerson et al.'s step-by-step and well-structured method for taxonomy development method [6] (see Table 2). This method has been frequently used in IS research [e.g., 17, 18]. As an input for Nickerson et al.'s ***empirical-to-conceptual*** (E2C) and ***conceptual-to-empirical*** (C2E) approaches, we review digital platform literature relying on [25][2] to not only derive dimensions and characteristics from extant research (C2E), but also to scrutinize digital platform instances studied in previous research to inform our taxonomy (E2C). These two approaches rest on our coding of 46 selected digital platform articles supported by *ATLAS.ti 8* as a technique in qualitative research to reduce data complexity [26].

Table 2. The Applied Steps of Nickerson et al.'s Method [6] in Our Taxonomy Development

Stage	*Stage's Application in Our Taxonomy Development*
1. Meta-characteristic: The meta-characteristic reflects the taxonomy's purpose that should rely on the taxonomy's expected use.	*Expected Use*: Digital platform designers, managers, and scholars seeking to classify digital platforms *Purpose*: Distinguish digital platforms based on their *high-level architectural configuration* (meta-characteristic)
2. Ending Conditions: Subjective and objective ending conditions determine when to terminate the method. Different ending conditions may generate different taxonomies.	*Objective Conditions*: The taxonomy consists of dimensions, each with mutually exclusive and collectively exhaustive characteristics. *Subjective Conditions*: The taxonomy must be concise, robust, comprehensive, extendible, and explanatory [6, p. 344].
3. Empirical-to-conceptual Approach: Reviewing a set of empirical instances (random, systematic, or convenience sample), the researcher tries to inductively group these instances' common characteristics into dimensions *without* considering existing conceptualizations.	*Sampling of Objects* (3.1): Coding of 46 selected papers yielding in a sample of 34 digital platform instances *Grouping of Objects* (3.2): Grouping of 34 digital platforms into 5 inductive, discriminate characteristics (*exchange*, *design* orientations; *direct*, *indirect*, *open* accesses) *Grouping of Characteristics* (3.3): Grouping of 5 characteristics into 2 inductive dimensions (*service* and *infrastructure* dimensions)
4. Conceptual-to-empirical Approach: Reviewing the previous taxonomy, the researcher tries to deductively conceptualize additional dimensions and characteristics that might not have been previously identified.	*Conceptualization* (4.1): Literature-based theorization of 2 additional deductive dimensions (*ecosystem* and *core* dimensions) *Examination of Objects* (4.2): Specification of 2 dimensions through 2 characteristics each (*private*, *federated* network; *access*, *resource* openness) after reviewing the sample of 34 digital platforms *Taxonomy Revision* (4.3): Revising final taxonomy (4 dimensions, 9 characteristics) to meet the ending conditions

[2] We search the *AIS Senior Scholars' Basket of Journals* in the *Business Source Premier* database employing the *EBSCOhost* search engine without time restriction. As digital platforms are an emergent modern concept, we also search the 2016/17 proceedings of ICIS and ECIS in the *AIS Electronic Library (AISeL)*. We select 16 journal papers, 10 ICIS papers, and 5 ECIS papers all of which have the phrase *"digital platform*"* in their abstract. The ISR special issue on digital platforms [14] is fully covered. A backward search adds another 15 papers.

First, we adopt the **E2C** approach in that we code the selected 46 papers to identify a sample of 34 digital platform instances that are studied in these 46 papers[3]. In the 1st iteration of the E2C approach, we randomly analyze 10 of the 34 instances (another 10 instances in the 2nd iteration; another 14 instances in the 3rd iteration). This resulted in extracting 5 inductive, distinct characteristics (*exchange*, *design* orientations; *direct*, *indirect*, *open* accesses). We group these 5 characteristics into 2 inductive dimensions (*service* and *infrastructure* dimensions) (see Table 2).

Second, we adopt the **C2E** approach in that we code the selected 46 papers to identify *existing* conceptions of digital platform characteristics. Therefore, we divide the 46 papers into 5 sets of 9, 9, 9, 9, and 10 papers, respectively. We thus embrace 5 iterations of the C2E approach in that we code 1 of the 5 sets of papers per iteration. We thereby identify 2 additional deductive dimensions (*ecosystem* and *core* dimensions). These 2 dimensions are specifically theorized in [3, 23]. In reviewing the sample of 34 digital platform instances, we further specify the *ecosystem* and *core* dimensions through 2 characteristics (*private*, *federated* network for the *ecosystem* dimension; *access*, *resource* openness for the *core* dimension). Rationalizing the overall 8 iterations in our taxonomy development, Table 2 synthesizes our methodological adoption of [6].

Digital Platform Archetypes. The next step is set out as the identification of digital platform archetypes. Therefore, we instantiate the taxonomy with the 34 digital platform instances to capture emergent archetypical configurations of digital platform profiles. Hence, we use architectural characteristics in each dimension as differentiating features of digital platforms to identify emerging dominant patterns that consistently reoccur. This is because digital platforms exhibiting *similar* characteristics along their dimensions should share identical architectural profiles, and belong to the same digital platform archetype [27]. We derive three dominant patterns of digital platform configuration as archetypes in this process. These archetypes are labeled as *orchestration*, *amalgamation*, and *innovation platforms* to reflect their main theoretical emphasis.

4 Taxonomy of Digital Platforms

Our findings suggest that digital platforms exhibit characteristics on at least four layered dimensions—namely, *infrastructure*, *core*, *ecosystem*, and *service* dimensions [e.g., 3, 23]. Afforded by the adopted platform architecture perspective, these dimensions reflect the socio-technical and complex architecture of digital platforms [1]. While the core dimension appreciates a set of stable technical core artefacts, the infrastructure, ecosystem, and service dimensions capture the dynamic periphery of platform components. Figure 2 sketches the identified *dimensions* that rest on the taxonomy's meta-characteristic to distinguish digital platforms from a platform architecture perspective. Figure 1 synthesizes each dimension's *characteristics* in a taxonomy of digital platforms in relation to Nickerson et al.'s *empirical-to-conceptual* (E2C) and *conceptual-to-empirical* (C2E) approaches (see Table 2).

[3] We list the 46 articles and the 34 digital platform instances in this database.

Architectural Dimension	Characteristic 1	Characteristic 2	Characteristic 3	E2C	C2E
Service	Exchange Orientation	Design Orientation	-	X	
Ecosystem	Private Network	Federated Network	-		X
Core	Access Openness	Resource Openness	-		X
Infrastructure	Direct Access	Indirect Access	Open Access	X	

Figure 1. Taxonomy of Digital Platforms

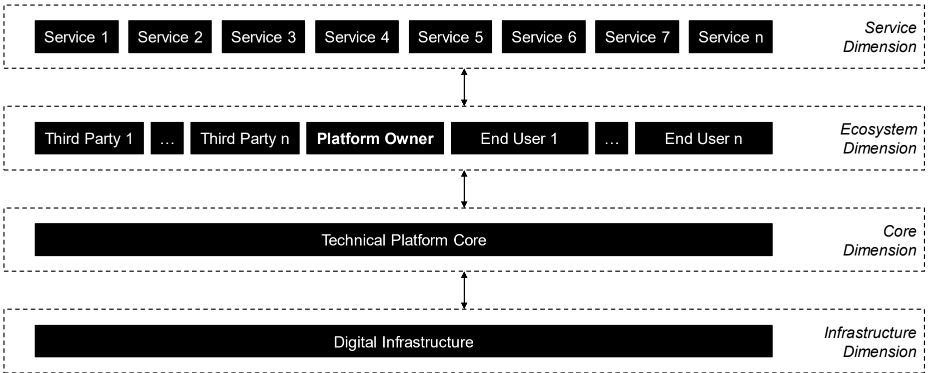


Figure 2. Digital Platforms' Dimensions from a Platform Architecture Perspective

4.1 Digital Platform's *Infrastructure Dimension*

Digital platforms are created and cultivated on top of **digital infrastructures**—here defined as *computing and network resources that allow distributed actors to facilitate their resource exchange* [5]. Examples of digital infrastructures include the Internet, data centers, open standards (e.g., IEEE 802.11 and USB), and consumer devices (e.g., smartphones and tablets). Digital infrastructures, therefore, are distinct from other types of infrastructures because of their ability to collect, store, and make digital data

Table 3. The Characteristics of Digital Platforms' *Infrastructure Dimension* [e.g., 5, 28]

Characteristic	*Definition*	*Rationale*
Direct Access [e.g., 28, 29]	Unobstructed access permission to an established digital infrastructure through the infrastructure owner that allows for guaranteed and instantaneous access	While reinforcing a platform's direct access rights through its enhanced status and market position, direct access infrastructures require costly access fees and extensive coordination between platform and infrastructure owners.
Indirect Access [e.g., 5, 28]	Obstructed access permission to an established digital infrastructure through intermediary, third-party access providers	Platforms with indirect access aim for hard-to-replicate partnerships with multiple intermediaries, allowing a plug-and-play strategy in selecting interchangeable intermediaries for cost reductions.
Open Access [e.g., 3, 29]	Unobstructed access to a new digital infrastructure devoid of permissions	Emulating direct access rights in a cost-effective fashion, open access infrastructures (e.g., blockchain) have lower market reach without comprehensive testing.

available across several systems and devices [28]. Relying on the E2C approach of our taxonomy development, we find that digital platforms access digital infrastructures in three ways—namely, *direct*, *indirect*, and *open* access. Table 3 outlines the *infrastructure dimension*'s three characteristics along with exemplary support from the literature.

4.2 Digital Platforms' *Core Dimension*

Digital platforms rely on a set of stable **technical core artefacts** of software and hardware. This set acts as technological foundation for a family of value-added platform derivatives [2]. These technical core artefacts denote an *extensible codebase serving as a building block upon which third parties devise platform-augmenting derivatives* (e.g., products, technologies, channels, and services) [16]. Relying on the C2E approach of our taxonomy development, we follow Karhu et al.'s [23] distinction of how core artefacts can interface with its periphery in two ways—namely, *access openness* and *resource openness* [23, pp. 3-6]—to promote third-party, platform-augmenting derivatives. Table 4 outlines these two characteristics of digital platforms' *core dimension*.

Table 4. The Characteristics of Digital Platforms' *Core Dimension* [e.g., 2, 16]

Characteristic	*Definition*	*Rationale*
Access Openness [e.g., 23, 30]	Granting of access to otherwise protected core artefacts to third parties by providing them with dedicated boundary resources to interact with the technical core artefact.	The rationale for *access openness* is to spark innovation within the platform ecosystem and induce third parties to use the core artefacts to create platform-augmenting derivatives that invoke positive network effects.
Resource Openness [e.g., 1, 23]	Opening the core artefacts' valuable resources by forfeiting their related intellectual property right (IPR)	The rationale for *resource openness* is that the technical core artefacts' owner sees it as advantageous to open the core resources by forfeiting related IPR.

4.3 Digital Platforms' *Ecosystem Dimension*

Digital platforms rely on a dynamic **platform ecosystem** here defined as *complex network of platform-mediated actor-to-actor interactions, turning increasingly accessible to end users through third parties' platform derivatives* [13]. Digital platforms are contingent on the availability and contribution of a critical mass of third parties within each of the relevant actor roles of the respective ecosystem. Prime examples for such actor roles are platform owner, partner, end user, and subcontractor [2, 31, 32]. Each of these actor roles offer complementary resources to the respective ecosystem to serve a wide range of end users and to satisfy various requirements [13]. For instance, Google generates most of its revenues within the Android ecosystem from advertisements powered through the use of its search engine, YouTube, and other Google services [23]. Relying on the conceptual-to-empirical approach of our taxonomy development, we follow Kazan et al.'s [3] distinction of two focal platform ecosystem characteristics—namely, *private network* and *federated network*. Table 5 outlines these two characteristics.

Table 5. The Characteristics of Digital Platforms' *Ecosystem Dimension* [e.g., 2, 13]

Characteristic	*Definition*	*Rationale*
Private Network [e.g., 1, 31]	Inward-looking, vertically integrated, and closed-loop ecosystem comprising an exclusive selection of private actors that shield their services from unauthorized actors	Private networks enact closed-loop systems to efficiently settle resource exchanges *within* their own boundaries. While the latter is virtually free, instantaneous, and guaranteed, resource exchanges *beyond* the closed-loop system demand fees, time, and risk from the private actors.
Federated Network [e.g., 1, 32]	Outward-looking, vertically disintegrated, and open-loop ecosystem mobilizing varied platform-augmenting third-party actors	Federated networks enact open-loop systems in which value creation and appropriation is distributed among federated third-party actors. These actors intentionally co-innovate with other external third-party actors to extend the capabilities and market reach of their mutual digital platform.

4.4 Digital Platforms' *Service Dimension*

With digital service being *the* value output of digital platforms [3], digital platforms eventually aim for and contribute to a gigantic shift from a product-based economy to one based on services, specifically *digital* services [22]. Digital service here refers to *an activity or benefit that at least one party can give to another, that is, predominantly provided through a platform-mediated digital transaction* [22, p. 507]. Notably, in contrast to classical bilateral owner-user relationships, platform-mediated digital service comprises a networked service *system* to integrate various organizational and technological resources to meet a given end user's needs. While the *giving* service offeror is the digital platform owner in cooperation with at least one platform partner, the *receiving* service beneficiary is the digital service user. Moreover, while a *single* transaction is sufficient to provide a digital service, often these transactions are provided *continuously* [33] and within actor-to-actor networks that configure the platform owner, at least one third party, and the end user in a unique manner [2]. Platform-mediated digital services are characterized by two distinct orientations—namely, *exchange* or *design* orientations. Table 6 outlines these two characteristics digital platforms' service dimension along with studies that support this dimension and its characteristics.

Table 6. The Characteristics of Digital Platforms' *Service Dimension* [e.g., 22, 33]

Characteristic	*Definition*	*Rationale*
Exchange Orientation [e.g., 34, 35]	Digital service aimed at reducing transaction costs in direct actor-to-actor exchanges	Exchange-oriented digital service (e.g., Facebook, PayPal, Uber, Airbnb) realize one-to-one matches between service offerors and beneficiaries and facilitate their subsequent direct exchange efficiently
Design Orientation [e.g., 1, 23]	Digital service aimed at enabling third parties to design platform derivatives and to disseminate them to a large audience	Design-oriented digital service (e.g., iOS, Android, Windows, Amazon Web Services, Linux) realize one-to-many matches between one third-party platform derivative designer and many derivative users

5 Archetypes of Digital Platforms

Our findings further suggest that digital platforms that exhibit *similar* characteristics belong to one of the three digital platform archetypes—namely, *orchestration*, *amalgamation*, and *innovation* platforms.

5.1 Orchestration Platform

Digital platforms that assemble federated networks—outward-looking, vertically disintegrated, and open-loop ecosystems—of platform-augmenting third parties through co-opetitive and inclusive platform profiles adhere to what we label as the *orchestration platform* archetype (see Figure 3). Orchestration platforms rely on high openness—both access or resource openness—to be highly integratable with existing third-party derivatives. These platforms' challenge is to derive a governance structure that *aligns* the business and technology interests among the platform owner and its many third parties. Orchestration platforms are highly dependent on established digital infrastructures (1) to connect third parties and end users; and (2) to attain elevated levels of joint market reach. However, each transaction on preexisting digital infrastructures negatively contributes to platform participants' costs as participants pay for access.

The expository case of *Android* [23] represents a prime example of orchestration platforms. The *Google*-sponsored open-source project (*access openness*) orchestrates a massive community of independent third-party developers (*federated network*) yielding in, depending on the estimate, a dominant 80%–90% share of the mobile phone market. Relying on *indirect access* to existing digital infrastructures (i.e., the Internet and mobile telecommunication infrastructures), its app store features over 3 million apps (*design orientation*) that generate more than 100 billion downloads per year [5].

Architectural Dimension	Characteristic 1	Characteristic 2	Characteristic 3	Digital Platform Archetype
Service	Exchange Orientation	***Design Orientation***	-	***ORCHESTRATION PLATFORM***: Co-opetitive and inclusive platform profiles
Ecosystem	Private Network	***Federated Network***	-	
Core	***Access Openness***	***Resource Openness***	-	
Infrastructure	***(Direct Access)***	***Indirect Access***	Open Access	

Figure 3. The *Orchestration Platform* Archetype

5.2 Amalgamation Platform

Digital platforms that assemble *private networks*—inward-looking, vertically integrated, and closed-loop ecosystems—comprising an exclusive selection of few private actors through monopolistic and assimilative platform profiles adhere to what we label as the *amalgamation platform* archetype (see Figure 4). Such platforms allow organizations to cultivate and grow private businesses without intervention from platform-augmenting third parties. In this sense, platform-mediated interactions are tightly controlled and directed inward to reinforce an insular digital platform. Amalgamation platforms are contingent on specific resources and capabilities to implement self-sustaining platforms by shielding their architectural dimensions from third parties. Moreover, such platforms are highly efficient, independent, and flexible in channeling their digital services through preexisting digital infrastructures. However, they face the challenge to maintain agility by avoiding the enactment of strategic linkages with third parties that are likely to introduce long-term legacy systems or platform derivatives. Such platforms rely on access to digital infrastructures to process digital services, while at the same time, seek to minimize resource outflows from its private network.

The expository case of *Pingit* [3] represents a prime example of amalgamation platforms. Launched by *Barclays* in 2012, this vertically integrated mobile payment platform captures value without third parties (*private network*). *Pingit* is designed to be a person-to-person (P2P) mobile payment exchange service (*exchange orientation*). Turning into a stand-alone application, however, it in turn incentivizes businesses to adopt *Pingit*. It is a proprietary mobile payment service as its development is fully internalized (restricted *access openness*). *Pingit* benefits from its direct access to *Faster Payments*, an existing digital infrastructure for mobile payments (*direct access*), to reach out to end users at rival banking institutions in a cost-efficient manner.

Architectural Dimension	Characteristic 1	Characteristic 2	Characteristic 3	Digital Platform Archetype
Service	***Exchange Orientation***	Design Orientation	-	***AMALGAMATION PLATFORM***: Monopolistic and assimilative platform profiles
Ecosystem	***Private Network***	Federated Network	-	
Core	***(Access Openness)***	Resource Openness	-	
Infrastructure	***Direct Access***	***(Indirect Access)***	Open Access	

Figure 4. The *Amalgamation Platform* Archetype

5.3 Innovation Platform

Digital platforms that assemble unobstructed access to a *novel* digital infrastructure devoid of permissions reverberate with our *innovation platform* archetype (see Figure 5). Such platforms embrace process innovation to deliver digital service through differentiated and cost-effective arrangements that are distinctively different from (and are seeking to transform) an industry's dominant process logic. This is realized through establishing—or forging strategic linkages with—*novel* digital infrastructures (e.g., blockchain). In this regard, affiliated stakeholders can circumvent the dominance of preexisting digital infrastructures—even though novel digital infrastructures that allow for open access may fail to become a dominant standard in facilitating digital services.

Architectural Dimension	Characteristic 1	Characteristic 2	Characteristic 3	Digital Platform Archetype
Service	***Exchange Orientation***	Design Orientation	-	***INNOVATION PLATFORM***: Hybrid and open platform profiles
Ecosystem	***(Private Network)***	***Federated Network***	-	
Core	***Access Openness***	***Resource Openness***	-	
Infrastructure	Direct Access	Indirect Access	***Open Access***	

Figure 5. The *Innovation Platform* Archetype

The expository case of *Blockchain.com* [3] leverages on third parties and subsidizing its digital services (i.e., payment, bitcoin wallets, exchange rates, JSON queries for blockchain data) for end users (*exchange orientation*). *Blockchain.info* thereby derives value from the bitcoin community by being integratable into various agnostic third-party services (*federated network*). Opening its core artefacts by forfeiting related IPR (*resource openness*), *Blockchain.info* operates on top of the *Bitcoin Blockchain*, an open digital infrastructure without access constraints (*open access*), to deliver bitcoins.

6 Discussion and Conclusions

Mediating various networked actors, digital platforms have become a pivotal means to shape digital ecosystems. We start with the premise that understanding and classifying digital platforms relies on a dedicated theoretical account on their *dimensions* and *characteristics* to postulate different configurations of digital platforms. Embracing the lens of platform architecture [7], we follow Nickerson et al.'s method for taxonomy development. The resultant taxonomy distinguishes digital platform instances through characteristics on their *infrastructure*, *core*, *ecosystem*, and *service* dimensions. We further disentangle *orchestration*, *amalgamation*, and *innovation* platform archetypes as a function of digital platforms' integral characteristics. We next discuss this study's theoretical and practical implications, limitations, and avenues for future research.

6.1 Implications for Theory and Practice

Theoretical Implications. *First*, we contribute to mitigating the outlined challenges of diversity in digital platforms' instances and conceptualizations. The 4 *dimensions*, 7 *characteristics*, and 3 *archetypes* serve as prospective theoretical means to more effectively guide and organize future theorization on digital platforms. That is, these means (1) seek to partially unify the variety of digital platform conceptions; and (2) classify digital platform instances. Thereby, through holding clearer definitions of *digital* platforms' dimensions and characteristics, the taxonomy considers the specific characteristics of digitality as an integral aspect of digital platform research in contrast to platform research in general [4]. We, thereby, hope to increase the comparability between digital platform instances and studies [5].

Second, we also contribute to mitigating the challenges of *vertically* and *horizontally* scoping digital platforms [1, p. 129]. *Vertical scoping* issues relate to choosing the appropriate level of the architecture for studying platforms. For instance, while mobile operating systems (e.g., Android) and associated app stores (e.g., Google Play) are often studied as the focal platform, new digital platforms are currently emerging *on top* of the mobile operating system (e.g., Facebook's Android app). To this end, our taxonomy contributes to disentangling the vertical scope of digital platform research through promoting four vertical architectural layers (see Figure 2). *Horizontal scoping* issues, in turn, refer to the variety of application domains (e.g., payment, health, banking, or mobile). Little research reflects the studied digital platform's application domain. The resultant lack of *contextualized* digital platform theory inhibits our understanding of how domain-specific digital platforms affect contextual outcomes. Characterizing digital platforms' context, the taxonomy facilitates more contextualized platform theory.

Practical Implications. This study contributes to the analysis of digital platforms [36] and is not prescriptive in nature. However, it has important implications for practice. *First*, without a well-developed taxonomy, it is difficult for managers and policymakers to differentiate diverse instances of digital platforms in their industries. Our taxonomy alerts these practitioners that not all digital platforms are equal, and it enables them to differentiate them regarding their characteristics, purposes, as well as required design decisions and institutional arrangements.

Second, this taxonomy becomes especially valuable when its characteristics and their distribution across digital platform instances are quantified. Specifically, practitioners can precisely measure various platform characteristics (e.g., types, frequencies, and durations of digital platforms' accesses to their underlying digital infrastructures) to link these measurements of diverse characteristics to digital platforms' differential impacts on platform outcomes (e.g., survival, performance).

Third, the paper identifies and differentiates three digital platform archetypes. These archetypes and their illustrative examples inform practitioners in making design decisions, as the archetypes represent ideal-typical configurations that have proven effective for platform survival and performance.

Fourth, the employed platform architecture perspective highlights the importance of developing the *architectural* approach to designing and maintaining digital platforms. Through embracing an architectural view on digital platforms, practitioners account for the socio-technical and complex nature of digital platforms [1]. Moreover, as the platform architecture perspective describes a digital platform's architectural configuration, practitioners are equipped to reflect the unique combination of their digital platform's components, respectively.

Fifth, policymakers can rely on the taxonomy in ensuring fair and efficient market regulations for digital platforms' constituent actors, which is in the interests of all platform participants, particularly in the light of lock-in and winner-takes-all effects. As such, the outlined dimensions and characteristics in the taxonomy inform policymakers in drafting legislative frameworks. Such taxonomy-informed frameworks would foster and regulate innovation effectively as the taxonomy allow policymakers to account for and balance multiple relevant dimensions of digital platforms. In turn, platform managers are provided with an organizing logic to more clearly define the specific aspects required in realizing *thriving* digital platforms. This may be especially useful for early design decisions that affect digital platforms' evolution trajectories. Managers might anticipate pivotal areas of concern and take appropriate measures.

6.2 Limitations and Future Research

While the taxonomy descriptively investigates digital platforms' dimensions, characteristics, and archetypes, it does not prescribe how to effectively configure digital platforms. Prospective research may thus investigate how different configurations translate into which outcomes (e.g., performance, survival, growth, flexibility, innovation). Further, we provide no statistical insights on what digital platform characteristics occur in which frequency. Future research may thus instantiate a *larger* sample of digital platforms to empirically learn more about the statistical distribution of characteristics.

Moreover, our taxonomy results from a restricted sample of 46 studies and 34 digital platform instances. Replicating our study with more digital platform instances in further contexts to validate and refine our taxonomy is thus pivotal. Moreover, our results are limited to the focused perspective of platform architecture. Alternative perspectives are likely to result in a different taxonomy and, therefore, in alternative archetypes. Therefore, due to opting for a specific perspective (i.e., platform architecture view), we neither claim exhaustiveness of the three derived archetypes, nor the comprehensiveness of the taxonomy in capturing all possible dimensions and characteristics.

Acknowledgements

This work has been supported by the Swiss National Science Foundation (SNSF).

References

1. Reuver, M.d., Sørensen, C., Basole, R.C.: The Digital Platform: A Research Agenda. Journal of Information Technology 33, 124-135 (2017)
2. Parker, G., Van Alstyne, M., Jiang, X.: Platform Ecosystems: How Developers Invert the Firm. MIS Quarterly 41, 255-266 (2017)
3. Kazan, E., Tan, C.-W., Lim, E.T.K., Sørensen, C., Damsgaard, J.: Disentangling Digital Platform Competition: The Case of UK Mobile Payment Platforms. Journal of Management Information Systems 35, 180-219 (2018)
4. Thomas, L.D.W., Autio, E., Gann, D.M.: Architectural Leverage: Putting Platforms in Context. Academy of Management Perspectives 28, 198-219 (2014)
5. Constantinides, P., Henfridsson, O., Parker, G.G.: Platforms and Infrastructures in the Digital Age. Information Systems Research 29, 381-400 (2018)
6. Nickerson, R.C., Varshney, U., Muntermann, J.: A Method for Taxonomy Development and Its Application in IS. European Journal of Information Systems 22, 336-359 (2013)
7. Cennamo, C., Ozalp, H., Kretschmer, T.: Platform Architecture and Quality Trade-offs of Multihoming Complements. Information Systems Research 29, 461-478 (2018)
8. Ciborra, C.U.: The Platform Organization: Recombining Strategies, Structures, and Surprises. Organanization Science 7, 103-118 (1996)
9. Gupta, A.K., Tesluk, P.E., Taylor, M.S.: Innovation at and across Multiple Levels of Analysis. Organanization Science 18, 885-897 (2007)
10. Yoo, Y., Boland Jr., R., Lyytinen, K., Majchrzak, A.: Organizing for Innovation in the Digitized World. Organization Science 18, 1398-1408 (2012)
11. Huang, P., Tafti, A., Mithas, S.: Platform Sponsor Investments and User Contributions in Knowledge Communities: The Role of Knowledge Seeding. MIS Quarterly 42, 213-240 (2018)
12. Lyytinen, K., Yoo, Y., Boland Jr., R.J.: Digital product innovation within four classes of innovation networks. Information Systems Journal 26, 47-75 (2016)
13. Wareham, J., Fox, P.B., Cano Giner, J.L.: Technology Ecosystem Governance. Organization Science 25, 1195-1215 (2014)
14. Sarker, S., Sarker, S., Sahaym, A.: Exploring Value Cocreation in Relationships Between an ERP Vendor and Its Partners: A Revelatory Case Study. MIS Quarterly 36, 317-338 (2012)

15. Lusch, R.F., Nambisan, S.: Service Innovation: A Service-Dominant Logic Perspective. MIS Quarterly 39, 155-176 (2015)
16. Gawer, A.: Platforms, Markets and Innovation. Edward Elgar Publishing (2011)
17. Siering, M., Clapham, B., Engel, O., Gomber, P.: A Taxonomy of Financial Market Manipulations. Journal of Information Technology 32, 251-269 (2017)
18. Prat, N., Comyn-Wattiau, I., Akoka, J.: A Taxonomy of Evaluation Methods for Information Systems Artifacts. Journal of Management Information Systems 32, 229-267 (2015)
19. Glass, R.L., Vessey, I.: Contemporary Application-domain Taxonomies. IEEE Software 12, 63-76 (1995)
20. Fiedler, K.D., Grover, V., Teng, J.T.C.: An Empirically Derived Taxonomy of Information Technology Structure and Its Relationship to Organizational Structure. Journal of Management Information Systems 13, 9-34 (1996)
21. Sabherwal, R., King, W.R.: An Empirical Taxonomy of the Decision-Making Processes Concerning Strategic Applications of Information Systems. Journal of Management Information Systems 11, 177-214 (1995)
22. Williams, K., Chatterjee, S., Rossi, M.: Design of Emerging Digital Services: A Taxonomy. European Journal of Information Systems 17, 505-517 (2008)
23. Karhu, K., Gustafsson, R., Lyytinen, K.: Exploiting and Defending Open Digital Platforms with Boundary Resources: Android's Five Platform Forks. Information Systems Research 29, 479-497 (2018)
24. Truex, D., Holmström, J., Keil, M.: Theorizing in Information Systems Research: A Reflexive Analysis of the Adaptation of Theory in information Systems Research. Journal of the Association for Information Systems 7, 797-821 (2006)
25. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, 13-23 (2002)
26. Holton, J.A.: The Coding Process and Its Challenges. Grounded Theory Review 9, 21-40 (2010)
27. Greenwood, R., Hinings, C.R.: Understanding Strategic Change: The Contribution of Archetypes. Academy of Management Journal 36, 1052-1081 (1993)
28. Henfridsson, O., Bygstad, B.: The Generative Mechanisms of Digital Infrastructure Evolution. MIS Quarterly 37, 907-931 (2013)
29. Tilson, D., Lyytinen, K., Sørensen, C.: Digital Infrastructures: The Missing IS Research Agenda. Information Systems Research 21, 748-759 (2010)
30. Ghazawneh, A., Henfridsson, O.: Balancing Platform Control and External Contribution in Third-party Development: The Boundary Resources Model. Information Systems Journal 23, 173-192 (2013)
31. Grover, V., Kohli, R.: Cocreating It Value: New Capabilities and Metrics for Multifirm Environments. MIS Quarterly 36, 225-232 (2012)
32. Tiwana, A.: Evolutionary Competition in Platform Ecosystems. Information Systems Research 26, 266-281 (2015)
33. Srivastava, S.C., Shainesh, G.: Bridging the Service Divide through Digitally Enabled Service Innovations: Evidence from Indian Healthcare Service Providers. MIS Quarterly 39, 245-267 (2015)
34. Bapna, R., Gupta, A., Rice, S., Sundararajan, A.: Trust and the Strength of Ties in Online Social Networks: An Exploratory Field Experiment. MIS Quarterly 41, 115-130 (2017)
35. Tan, B., Pan, S.L., Xianghua, L., Lihua, H.: The Role of IS Capabilities in the Development of Multi-Sided Platforms: The Digital Ecosystem Strategy of Alibaba.com. Journal of the Association for Information Systems 16, 248-280 (2015)
36. Gregor, S.: The Nature of Theory in IS. MIS Quarterly 30, 611-642 (2006)

Value of Star Players in the Digital Age

Christoph Buck[1], Sebastian Ifland[1], Michael Renz[2]

[1] University Bayreuth, Information Systems Management, Bayreuth, Germany
{christoph.buck,sebastian.ifland}@uni-bayreuth.de
[2] University Bayreuth, Sport Governance and Event Management, Bayreuth, Germany
{michael.renz}@uni-bayreuth.de

Abstract. International professional football has become a billion dollar market worldwide. Up to half of the world's inhabitants watch major events such as the FIFA World Cup or the UEFA Champions League Final. Central players and elements in this global advertising market are the clubs' teams and the players themselves. Social media platforms today allow professional footballers to reach millions of people through private marketing as individuals, thereby creating their own brand. The brand and reach of each player is also a valuable resource for clubs in terms of player value and transfer activity. The outlined results in the following article show indicate a positive correlation between the social media value of professional players and the transfer activities of football clubs. Consequently, the impact of digitization on professional football can be shown by a relationship which has not been investigated in research to date.

Keywords: social media, transfer market, digital strategy, football, star marketing

1 Introduction

Football is the world´s most popular sport and is no longer perceived solely as a game but also as a battle for financial success [1]. The Fédération Internationale de Football Association (FIFA) alone generated sales of USD 734 million in 2017 [2]. The development and growth of the industry reflects the immense interest of the sport spectators [3]. Half of the entire human population is enthusiastic about football; 3.2 billion people alone watched the 2014 FIFA World Cup [4]. No other industry in the world has such a level of attention and awareness.

Due to this level of attention, it would seem natural that the best clubs and players themselves create strong brands. Nowadays, the most important information systems for these actors are social media (SM) platforms like Twitter, Instagram, and Facebook [5]. Football companies and the players have a unique advantage over all other companies when it comes to SM. Fans are intrinsically motivated to consume the content provided by clubs, while traditional companies must actively contact customers to attract their attention [6]. The only challenge for the sport’s clubs is to attract fans from other clubs. An effective method is obvious: star marketing [7]. Fans are interested in clubs on the one hand, and in the players themselves on the other. A logical

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

conclusion would be to oblige all players to enhance their reputation among fans and thus be valuable for the club's SM appearance. However, to what extent the clubs have acted in this way has not been investigated in any field of research to date.

A lot of work was done on the use and value of SM for sports clubs [1,7,8]. However, to date, there is a research gap relating to the way and extent the SM value of players shape clubs' SM reach and how this is taken into a club's consideration regarding actual player transfer activities.

Consequently, in this paper the relationship between the SM value of players and related transfer activities will be examined. It is assumed that the purchase of a player on the one hand is driven by his value as a football player, and on the other hand also by his brand value. Consequently, transfer activities might be impacted by the SM value of the players, as SM is the predominantly used information system for brand development in the market. We provide a model to investigate this relationship. Concretely, the following research question should be answered:

Is there a correlation between SM parameters and transfer activities, and could professional clubs take this into consideration?

To answer this research question, this article is structured as follows: In the following section the theoretical foundations and relevant work of success criteria and brand management in sports are discussed. In the next section we apply the chosen research method to the field. Subsequently, we present the results and discuss the main findings. Finally, a conclusion is provided containing limitations, implications and future research.

2 Theoretical Background

2.1 Success Factor Player

First of all, in terms of market theory, the difference in a sporting and an economic competition between football clubs has to be remarked upon [9,10]. Sporting competition is characterized by sporting success as the defined market outcome. In relation to one's opponents, those who achieve better placings in competitions are successful in sport [11]. With respect to sporting outcome, the acquired player talent is the most influential determinant in football clubs' production functions [12-14]. The distribution of player talent therefore shapes the sporting competition and determines sporting success in team sports. The success in sport competitions, in turn, is of central importance for the economic success of clubs, defined as the greatest possible outcome on product markets and thus generation of revenue streams (e.g. tickets, media rights, merchandise items) [15]. In a simplified scheme, sporting success causes higher demand and results in more trophy money [16]. Team sports contain attributes of a positional good where marginal changes in sporting success cause huge differences in revenues [17-19]. However, players' talent or sporting success is only one determinant for the revenues generated on product markets. In addition, an impact is made by market size, consumer purchasing power, substitution possibilities, club history, fan culture, uncertainty of outcome or the team quality, including "superstar effects" [16,20].

However, sporting success not only influences economic success, but is also in turn highly affected by the economic potential the other way around. Economists describe this intertwining by the concepts of derived demand and marginal revenue product [21-23]. In simple terms, the marginal revenue product (MRP) is referred to as the marginal productivity of the player's labour effort in relation to the revenue the club can generate from that marginal product [24-26]. The higher the club´s revenue streams or the player`s ability to increase revenue streams, the higher the player´s MRP. Consequently, a higher MRP implies a higher demand for this player leading, ceteris paribus, to higher wages and transfer fees. In sum, the MRP is determined by club-related economic criteria as well as player-related characteristics. Big market clubs manage to generate more money regardless of the acquired playing talent. Consequently, these clubs are willing to pay higher wages and transfer fees caused by a generally higher revenue and MRP level.

Turning to the players' side, MRP relevant factors can be distinguished in two aspects [27]. The first part describes the players' contribution to sporting success and related revenue, including trophy money. Furthermore, and besides actual on-pitch performance, a player-related revenue could be generated because of his image and brand value. Hence, the MRP in a single period can be specified as [27] :

$$MRP = (PWC \times MWR) + PBV \tag{1}$$

PWC stands for player win contribution or, in other words, the marginal gain in sporting team success. MWR is the marginal win revenue the club can generate out of this marginal gain in sporting success. PBV means player brand value and represents the brand-related revenues which can be made out of the additional media and spectator attraction.

The effect of generating additional and often disproportionate revenues solely due to the image or brand of a player is often referred to superstar effects [28,29].

As sporting competition is mainly driven by player talent, and, as seen above, talent demand is strongly connected to economic demand, it appears that economic competition would seem to highly correlate with sporting competition [30]. Obviously, from both an economic and a sporting point of view, players are the clubs' most important investment [31].

The transfer markets are the central platform to invest in players and to process player transfers. Transfers are often based on different MRPs indicating different demand curves between buying and selling clubs. Greater MRP differences due to greater revenue divergences trigger player transfers and are clearly among the main reasons for expanding player migration in global football, although if a player is to change clubs before the end of his contract, this contract must be terminated. The buying club pays the selling club a transfer fee so that the seller is willing to terminate the current contract with the player [32]. In relation to the Union des Associations Européennes de Football (UEFA), the market value of a professional athlete represents the monetary amount the club is willing to pay in order to contractually bind this athlete [33]. The market value of all players in a club results in a total team market value [34]. As seen above, the market value of players expresses the sum of the discounted MRPs over the length of the player`s contract [27]. On the premise that clubs act with economic rationality, transfer fees should not exceed the MRPs and should reflect the market value.

2.2 Social Media Brand Value

Advertising and transmission fees represent the largest sources of income for clubs [35]. The strategy of owning a brand strong enough to generate sufficient revenue from merchandise, ticket sales and international transmission is therefore a key objective in team sports economic competition [11]. A brand is the focus of all expectations and opinions of customers, employees and other stakeholders about an organization and its products and services. According to Brand Finance, a brand's technical implementation comprises the components contribution, strength and value [36].

Television has changed sports brand marketing and the sports information industry. Similarly, SM also triggered a great change [7], which brings with it the digital transformation of sports companies [37]. A strategic orientation towards SM is necessary for the successful brand management of football clubs [37]. One reason for this is that, more than ever, brands are required to find new ways to communicate with young consumers who spend less time on television, print and other traditional media, and are more likely to be addressed by digital marketing communication [38]. SM is increasingly seen in this context as an additional marketing channel through which companies can communicate and interact with their consumers or potential customers [39]. SM marketing is therefore primarily about listening to the users of SM and responding appropriately [40]. These dialogues have a direct positive effect on the advertisers' results by driving the fans' behavior towards the consumption of the advertised services and products [41]. Professional sport associations are trying to address and monetize this potential use of SM marketing [8]. This is currently done via SM communication [8] and Web 2.0 technologies [40].

The SM presence is subject to a complex composition, though. The presence refers to all engagements [8] which are implemented with club-own or player-own profiles in social networks. In addition, it is important to implement a purposeful approach to the player's brand, in order to distinguish him as an independent brand from the mass of professional football players. As part of sports marketing, this refers to star marketing and could be of great importance for the brand of a club, because the combination of different SM profiles has a positive effect on the interaction of each individual page [7]. Digital reaches are created by the profiles of the clubs on SM channels [8] and are reinforced by profiles of the players, because each player represents the image of the club [42].

The overall aim of this paper is to disclose player transfer effects and the inherent player SM activities on the clubs' SM appearance and to interpret the influence of SM on the transfer activities itself.

3 Research approach

3.1 Model Assumptions

To design the relationship, the complex reality is broken down into a simplified model. The following axioms therefore apply to the further procedure, which are intended to illustrate complex real interrelationships in a simplified manner:

1. A search query of an individual (fan) to the search engine Google implies interest of this individual (fan) in the object (transfer or club) or subject (player) searched for.
2. The interest of an individual in an object (transfer) or subject (player) lies in the expected change of his state (transfer of player).
3. The interest of an individual (fan) in the expected change of state of a subject (player) justifies the willingness to use information channels (SM channels of the clubs).

Football fans exist as supporters of clubs and as supporters of individual players. Only negative associations can arise, therefore, when players leave. Although there are cases in which player changes among local rival clubs lead to a negative mood among fans, these are neglected in the present analysis. On the one hand, no profound rivalries between the old club and the new club could be found in any of the transfers levied. On the other hand, only new players are considered. From this follows the continuing formulation of two further axioms:

4. A change in the state (change of players) of a subject (player) can only positively influence the information channels (SM channels of the clubs) of an object (club).
5. The regularity of exceptional events confirms their dependence on a regularly recurring variable.

Based on these assumptions, the procedure is further divided into three steps. The first section deals with the study of the existence of a stochastic dependence of the growth of the clubs' SM reach in the season, particularly the transfer period (Section 3.2). The second step aims to identify and describe the direct influence of a player transfer on the SM accounts of the clubs (Section 3.3). Lastly, the final step is to categorize transfers to obtain an assessment of club activities (Section 3.4).

3.2 Stochastic Dependence

If fans have a special interest in player changes, that interest will affect the search hits and therefore affect the clubs´ SM reach regarding to the above-mentioned axioms 1 & 2. The effect we are looking for lies in the distribution of clubs´ SM reach growth peaks. Players can only be transferred during those periods and hence a fans´ action on interest in player changes appears during transfer periods. Related to this constraint, growth peaks of clubs´ SM must depend on the transfer periods. To prove this, we investigate the stochastic dependence of both quantities. Stochastic independence from events means that they do not influence each other or are influenced jointly by other events. This relationship makes it possible to determine the direction of an existing stochastic dependence from the assumed causal relationship. The existence of stochastic independence is confirmed if formula 2 is fulfilled.

$$P(S \cap T) = P(S) \times P(T) \tag{2}$$

The existence of stochastic dependence is therefore confirmed in the event of inequality between the two terms. The variables S or T stand for event one or event two. Event one (S) represents the absolute count of growth peaks. Event two (T) therefore outlines the absolute count of data points within the transfer period. The data collected are available in the form of absolute daily growth. Each data point consists of both the index T whether it is inside the transfer period or not, and the index S whether it is a growth peak or not. Growth peaks S are extracted using boxplot diagrams. Then the distribution of the outliers over four intervals was measured.

3.3 Interest in Players Analysis

The aim of this paper is to investigate the possible impact of SM on the transfer activities of professional sports clubs. With this in mind a special focus lies in asking to what extent player changes and therefore the interest of fans on players can influence the SM reach of sport clubs. If an effect of a transfer on the clubs´ SM channels can be shown, a new dimension of players´ value will be identified. This digital transformation-driven value might open up new types of transfer fee amortization and income.

To analyze this influence, we focused on the growth behavior of the SM reach of the clubs during the weeks of the year in which transferred players received the highest media interest. Doing so allows us to extract the periods in which the player was of special interest among fans because they looked for information about him. Corresponding to the above-mentioned axioms 2-4, the fans´ interest causes SM engagement with the clubs´ channels. In contrast to the method used in Section 3.2, which was used to identify the link of SM growth and the transfer periods, the approach described in this chapter quantifies the direct influence (E_{dir}) of transfers on the SM channels and therefore the willingness to pay for players with regard to their digital value.

The analysis tool Google Trends is highly valuable in recognizing fans' interest. Using this tool allows us to identify the search relevance per week of a player on the search engine Google over the last 12 months [31]. Again, regarding axiom 3, this exposed interest can be related to the willingness of the searcher to get information about a player and therefore encourages the usage of information channels about the player. Consequently, an increase in hits to players´ and clubs´ SM channels is predicted. The method used in this chapter is meant to show a direct influence (E_{dir}) which would prove the mentioned relation between fans´ interest and an increase in SM channels´ followers.

$$E_{dir} = Z_{middle,week} - Z_{middle,period} \tag{3}$$

The variable $Z_{middle,week}$ characterizes the mean daily SM follower growth during the week within the considered period in which the search volume of an individual was of highest interest on Google´s search engine. In contrast, $Z_{middle,period}$ stands for the mean follower growth of the individual´s channels over the entire survey period. The difference between the average daily increase within the week in question and the average daily increase over the survey period represents the direct effect E_{dir} of the player transfer on the clubs' platforms. The search peak of the new player and not the date of the change of player was selected as the period to be examined. This made it possible to include the reaction speed of the interested individual. The premise was that a fan can only become active on search engines and social channels as soon as he or she learns of at least one rumor about a possible transfer or the execution of the transfer itself.

3.4 Transfer Classification

After examining the direct influence of the transfers on the channels of the clubs, the individual player changes were categorized. For this we took into consideration four basic characteristics: (1) new player improves the starting line-up, (2) new player uses SM, (3) new player causes a recognizable effect (E_{dir}) on clubs´ SM channels´ followers, and (4) new player evinces SM potential. All of the characteristics have been regarded as dummy variables. The changes were hierarchically assigned to the classes: supplementary player, SM reach applied player, balanced player, sporting improvement player, SM star player and star player.

Previous research took various factors such as fitness, minutes played, and duel strength into account to compare sporting analysis units [15,31]. For this study, however, the sporting performance of a player was limited to the achieved point average per game and the average minutes played per game. A player who plays for a team that has a high average score and is therefore successful shows a high sporting quality simply through his status as a team member. The inclusion of the average minutes played per game additionally specifies the qualitative classification of the player. The importance of the player in the team is derived from this, since better players are usually used more frequently [31].

To measure the sporting performance, a new variable F was implemented. As mentioned above, F consists of the average minutes played and the achieved point average during the season 2016/17 ($PPS_{16/17}$). Formula 4 shows that the average minutes played per match were calculated by dividing the absolute minutes played by the player during the 2016/17 season ($M_{16/17}$) by the total number of matches played by the team the player represented during the 2016/2017 season ($G_{16/17}$).

$$F = M_{16/17} * (G_{16/17})\text{-}1 * PPS_{16/17} \quad (4)$$

If a player has a better score than his worst position-related competitor, this transfer was considered a sporting improvement. Not all players of the squad were used as a yardstick: only the players of the starting line-up. The starting line-up is based on a 4-4-2 system as standard. A midfielder was therefore a sporting improvement if at least one of the four best midfielders has had a worse factor than the player in question.

The attributes 'uses SM' and 'effect recognizable' reflect the ability of the player to gain a digital advantage for the buying team. The former attribute includes the player's own digital commitment. A positive manifestation exists as soon as the player in question operates one or more official SM channels. The latter feature, on the other hand, is based on the findings of the above-mentioned interest in player analysis. This attribute is available as soon as a positive effect has been recorded ($E_{dir}>0$).

The SM-sports weight ratio (WR_{SMS}) ultimately provided information about the transfer motive.

$$WR_{SMS} = (\text{SM reach}) * (F)^{-1} \quad (5)$$

As a quotient of the SM reach and the player's sporting performance factor, the WR_{SMS} stands for the SM potential of the player. An above- or below-average ratio indicates a large (small) or small (large) digital (sporting) potential. All above-average digital potentials that could be identified as outliers via a boxplot diagram were classified as exceptionally strong. A weak potential is automatically present if the player in question does not have a SM channel. It is important to note that we did not investigate where the player's followers came from, for which reason they became followers, or who manages the profile.

The classification was executed using a Bayes Classifier. Therefore, we provided the above-mentioned characteristics to a sample data set. We then applied a Bayes learner as shown in Figure 1. The sample data used represented all combinations of states that the variables could assume.

The learner then predicted the algorithm learned from the sample data for the Bayes Classifier which we executed on the main data set of 124 player changes. As a result, every transfer was matched to one of the five classes described in the following: A balanced player may use SM and possibly might have a direct effect, but the sporting factor must be weighted more strongly in relation to its digital reach. Therefore, this player is neither an enormous sporting improvement nor an outstanding digital improvement. A sports-related transfer is therefore every player who represents an improvement in sports and whose weighting tends to be in addition to the sporting value. If there was a sporting improvement together with a strong or extraordinarily strong SM potential, this player would represent a star transfer. Conversely, SM stars have an exceptionally strong SM potential but cannot remarkably improve the squad in sporting terms. A player with a higher weighting on the reach, who either cannot have a direct effect or does not represent a sporting improvement, is assigned to the reach class. Finally, all players who do not fulfill any of the attributes are additional players.

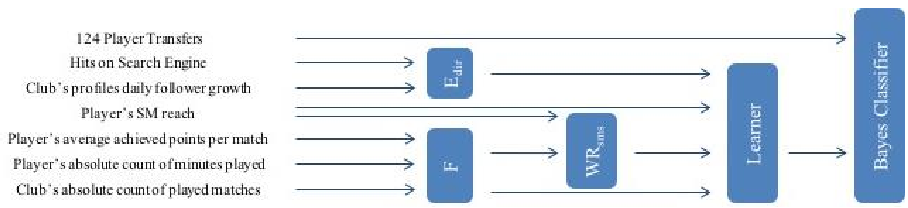


Figure 1. Statistical Approach

3.5 Data Collection

The analysis focuses on the 20 clubs with the highest income in football worldwide [35]. Müller et al. [31] point out that in professional football, the channels YouTube, Facebook, Twitter and Instagram are the most important. YouTube is a video portal and primarily serves to support the sporting event [40]. For this reason, it is assumed that growing reach in this medium is based more on sporting highlights than on transfer effects. The analysis used daily followers growth from all official channels of the clubs on Twitter and Instagram platforms. Channels which obviously had no relevance for the men's football team (youth, women's division, handball, etc.) were ignored. A total of 43 official Twitter and 20 official Instagram channels were identified. The historical data was collected by the online platform SocialBlade, which resulted in a dependence on the platform in terms of data quality and completeness. The study took place from November to December 2017. The increases were considered in the period from May 1, 2014, to October 31, 2017. Missing data entries were added using linear interpolation.

In addition, data on transfer transactions were used. Due to the lack of coverage of historical data on Twitter profiles, the individual consideration of the transfers had to be limited to the summer change period of the 2017/18 season. The Transfermarkt.de platform, which is characterized by high data quality and domain-related relevance [31], collected all new transfer entries for the aforementioned change period of all 20 clubs considered. All returns on loan that had already been purchased and subsequently borrowed in previous periods were excluded. Also disregarded were former youth players who transferred from the youth team to the men's team of the same club. The survey yielded 124 hits. For the all the relevant transfers found, the existing SM reaches of the players on Instagram and Twitter channels were collected on May 30, 2017. The SM reach represents the sum of the followers on both channels. The weekly time period for each transfer in which the search volume of the player concerned was highest was then collected with the help of Google Trends. The period in which the maximum number of search queries was determined was twelve months retroactively from December 31, 2017. From a total of 37,868 data points, 1,904 profile-related outliers could be identified as growth peaks S. As expected, 30.79% of all data points were within the transfer period due to the seasonal distribution. 73.58% of the growth peaks were in the period of one transfer period.

4 Findings

The application of formula 1 to the values resulted in an inequality. This result confirms the existence of a stochastic dependence of the variable peak growth and transfer period. As stochastic dependence indicates only the existence of a correlation and not the cause of it; the direction of this dependency cannot be determined from the numerical values. Nevertheless, 73.58% of all growth peaks were in the transfer period interval and 70.97% of the considered transferred players experienced the largest search volume in at least one week which intersects the transfer period.

For the purposes of this review, a positive effect of a transfer for a club has been defined as an impact on its SM commitment. In this model, this effect was assigned to the growth of the follower numbers. From the connection of the week of the search peak and the increase of the SM channels, the key figure of the direct effect could be calculated. Table 1 shows the effects of the summer transfer period 2017/18 and the transfer sums for each player.

91 out of 124 newcomers use SM. At the same time, 72.53% of these were able to gain a direct effect of an average of about 4.5 thousand new followers. The medium SM reach of these players thus exceeded the medium reach of players without effect by about 2.8 million followers. The average transfer fee for players with an SM appearance was about twice as high as for those players who do not maintain a presence in social networks.

Table 1. Transfers and SM usage

Influence	Descriptive statistics	Uses SM	Doesn't use SM	Total
$E_{dir} > 0$	Number	66	19	85
	Share (within vertical category)	72.53%	57.58%	
	Share (within horizontal category)	77.65%	22.35%	
	Share (of the whole)	53.22%	15.32%	
	Average transfer fee (€)	25,093,878.79	10,036,842.11	
	Mean positive effect (adjusted for mean growth)	4,472.83	3,859.21	
	Average follower number	3,498,609.42	-	
	Correlation follower and E_{dir}	0.2988	-	
$E_{dir} \leq 0$	Number	25	14	39
	Share (within vertical category)	27.47%	42.42%	
	Share (within horizontal category	64.10%	35.90%	
	Share (of the whole)	20.16%	11.29%	
	Average transfer fee (€)	13,623,800.00	6,230,769.23	
	Average follower number	726,469.12	-	
Total		91	33	124

The data also show that 57.58% of the transferred players without a SM appearance also have direct positive effects. The transfer sum for those players without SM channels is also approaching at least the transfer sum for players with a SM presence. The average transfer fee paid for players without SM and with effect was closer to the average amount paid for players with SM and without effect than for players without SM and without effect.

Finally, Table 2 presents the classification of individual transfers for the 2017 summer change period. 6.45% of the transferred players were purchased for an average of 5.75 million euros. Balanced players make up a share of 29.84% with an average transfer fee of 11.8 million euros. With 33.87%, the sports class represents the largest share of transfers. The average transfer fee is 17 million euros above the average for additional players, but below the sums paid for the individual classes with SM motives. The three categories that indicate a SM motive (reach, SM star, star) account for a total of 29.83%. Consequently, almost every third transfer could at least be associated with SM strategic motives.

Table 2. Classification

Class	Number	Share	Average transfer fee (€)
Additional player	8	6.45%	5,750,000.00
Balanced player	37	29.84%	11,841,756.76
Sport improvement	42	33.87%	17,084,047.00
SM reach improvement	10	8.06%	18,615,000.00
SM star	10	8.06%	19,310,000.00
Star player	17	13.71%	41,382,352.94

5 Interpretation

The results show that the development of the SM reach of clubs is not a coincidence. As seen in the findings, the greatest growth spurts result from events that occur within the transfer window, hence a concrete connection between the SM growth and the annual transfer period was confirmed. Due to the given assumptions, this indicates that transfers have an impact. Being able to achieve a SM effect through a transfer therefore should influence the transfer decision itself. This explains SM's influence on the transfer decision and thus on the transfer market, because clubs try to use SM as a strategic means to ensure long-term success [8,11].

However, there were also growth peaks outside the transfer period. These peaks can be explained either by sporting events, i.e. particularly sensational events, or speculations in the run-up to an imminent transfer [5]. However, since these are the minority, the question of a connection between the appearance of players and clubs on social networks and the transfer market can be answered affirmatively.

Regarding transfers in the context of this paper, the SM motive and therefore the influence of SM on the transfer market is associated with an additional reach, which is transferred to the sphere of influence of the club in the form of the reach of the new player. This might influence the willingness of clubs to pay for players, as the higher transfer sums indicate. This expresses the potential benefit via the indirect digital expansion through followers of the player profiles. The potential can be explained by the fact that fans and followers of the player can identify more strongly with a club if the favored player is under contract with the club. The media interest of the fans in the club increases through the interest in the player.

An unexpected direct effect was caused by players who do not have a SM profile. A possible explanation for this unexpected finding is once again the interest of fans in re-accessing the game. However, this creates a conflict, as consumers have no direct access to the player due to the player's lack of digital presence. Clubs can integrate these players into the generated content as part of the star marketing effort [7].

Due to the greater transfer sums, it can be assumed within the framework of the assumptions made that the clubs are aware of the digital potential from transfer transactions. Nevertheless, the values indicate a lack of knowledge in dealing with the digital potential of the players. The potential benefit consists of the components ‘direct effect’ and ‘additional reach’ through player profiles. The results suggest that the clubs are only aware of the players who have SM profiles on their own. The average fee paid for players who don´t have SM but could impact a direct effect are lower than the fee paid for players who have SM but could not impact a direct influence. Therefore, if clubs would have been aware of those hidden players they would have been transferred with higher fees, especially when recognizing the average follower count of the players that do have SM but could not cause a direct effect being about 700,000. That volume of followers seems not to justify paying a higher fee.

Lastly, the differentiation between the six classes reflects the significance of the SM component in the transfer market. It was to be expected that star newcomers would require the highest transfer fees. However, it is noticeable that both SM stars and purchases for the reach are generally more expensive than for sporting improvements. This means that a player who only offers added value in the digital field is of greater value for sports clubs than a player who only contributes on the field. Against the background of any commercialization and globalization of sport, it can be confirmed within the framework of the assumptions made that it is no longer athletics but brand value that is in the foreground.

6 Conclusion, Limitation and Further Research

The model and the analysis steps depict reality in a highly simplified manner. Although it was suitable for identifying anomalies, the model does not manage to explain these anomalies in their complexity of real circumstances. This requires a much more specific and precise consideration. Both the categorization of transfers into classes and the comparative classification of club strategies offer a good approach for further refining the transfer analysis and still providing more sustainable models for the analysis of modern football than are available today [15]. Moreover, a more detailed analysis on individual players (e.g., good players in worse clubs) should be added to strengthen the model and analysis. Nevertheless, it is necessary to extend the transfer parameter set in subsequent analyses in order to take club and sporting MRP components into account. In addition, the sustainability of the analysis should be strengthened by extending an analysis to other SM platforms. Müller et al. [31] and Buhler et al. [40] cite as examples platforms which are of the greatest importance in professional association football. A local analysis is another component that should be added to the model developed in order to identify the strategic actions of clubs. Finally, it is recommended to extend the analysis to a randomized sample in order to cover the entire global professional association football.

This study should be seen as a first step in the field. The authors tried to operationalize the underlying assumptions. As the results show, there is a relationship between the SM value of players and the transfer activities of clubs. Further research should develop the model to a more theory-driven basis and incorporate more parameters and a greater sample size.

As professional sports and SM do have a huge impact on social life, both academia and practice should increase the research in this field. Furthermore, professional Football Clubs should take this topic more seriously. As could be seen in the data, there are informational gaps about the potential value a player can add to the club by using SM. The teams should strengthen their knowledge in identifying those players and their potential values in a digitalized society, especially those players without SM profiles on their own.

References

1. McCarthy, J., Rowley, J., Ashworth, C.J., Pioch, E.: Managing brand presence through social media: the case of UK football clubs, Internet Research 24 (2), pp. 181–204 (2014)
2. FIFA, Activity Report (2017)
3. Riedl, L.: Vom Spiel zum Spektakel. bpb (2016)
4. FIFA, https://de.fifa.com/worldcup/news/fifa-fussball-wm-2014 (Accessed: 16.12.2017)
5. Chen, H., De, P., Hu, Y.J.: IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales, Information Systems Research 26 (3), pp. 513–531 (2015)
6. Boyd, D. M., Ellison, N.B. Social Network Sites: Definition, History, and Scholarship, Journal of Computer-Mediated Communication 13 (1), pp. 210–230 (2007)
7. Holland, C.P.: Internet and Social Media Strategy in Sports Marketing (2015)
8. Wulf, J., Söllner, M., Leimeister, J.M., Brenner, W.: FC Bayern München goes Social. ECIS 2015 completed reseaech papers, AIS Electronic Library (2015)
9. Dietl, H.M., Franck, E., Hasan, T.: Zur Eignung der Ligagenossenschaft als Organisationsform im professionellen Teamsport — Eine modelltheoretische Untersuchung, Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 59 (4), pp. 447–464 (2007)
10. Schellhaaß, H.-M., Enderle, G.: Die zentrale Vermarktung von Europapokalspielen aus ökonomischer Sicht, Wirtschaftsdienst 78 (5), pp. 294–300 (1998)
11. Schilhaneck, M.: Zielorientiertes Management von Fußballunternehmen. Gabler Edition Wissenschaft (2008)
12. Swieter, D.: Eine ökonomische Analyse der Fußball-Bundesliga. Duncker & Humblot, Berlin, (2002)
13. Teichmann, K.: Strategie und Erfolg von Fußballunternehmen. DUV Deutscher Universitäts-Verlag, Wiesbaden (2007)
14. Downward, P., Dawson, A., Dejonghe, T.: Sports economics: Theory, evidence and policy, Butterworth-Heinemann, Amsterdam, London (2009)

15. Ruijg, J., van Ophem, H.: Determinants of football transfers, Applied Economics Letters 22 (1), pp. 12–19(2015)
16. Borland, J., MacDonald, R.: Demand for Sport, Oxford Review of Economic Policy 19 (4), pp. 478–502 (2003)
17. Thieme, L.: Zur Konstitution des Sportmanagements als Betriebswirtschaftslehre des Sports: Entwicklung eines Forschungsprogramms. Zugl.: Saarbrücken, Univ., Habil.-Schr, Epubli, Berlin (2011)
18. Kruse, J.: Ökonomische Konsequenzen des Spitzensports im öffentlich-rechtlichen und im privaten Fernsehen, Helmut-Schmidt-Universität der Bundeswehr Hamburg, Hamburg
19. Vöpel, H.: Do we really need Financial Fair Play in European club football? An economic analysis, ifo DICE Report 9 (3), pp. 54–59 (2011)
20. Budzinski, O., Feddersen, A.: Grundlagen der Sportnachfrage: Theorie und Empirie der Einflussfaktoren auf die Zuschauernachfrage (2015)
21. Scully, G.W.: Pay and performance in major league baseball, The American Economic Review 64 (6), pp. 915–930 (1974)
22. Neale, W.C.: The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition, The Quarterly Journal of Economics 78 (1), pp. 1–14 (1964)
23. Berri, D.J., Leeds, M.A., Allmen, P.v.: Salary Determination in the Presence of Fixed Revenues, International Journal of Sport Finance 10 (1), pp. 5–25 (2015)
24. Rosen, S., Sanderson, A.: Labour Markets in Professional Sports, The Economic Journal 111 (469), pp. 47–68 (2001)
25. Scully, G.W.: Pay and performance in major league baseball, The American Economic Review 64 (6), pp. 915–930 (1974)
26. Gerrard, B.: Achieving transactional efficiency in professional team sports: the theory of and practice of player evaluation, in: J.A. Goddard, P.J. Sloane (Eds.), Handbook on the economics of professional football, Edward Elgar, Cheltenham, UK, pp. 189–202 (2014)
27. Nüesch, S.: The economics of superstars and celebrities, DUV Deutscher Universitäts-Verlag, Wiesbaden (2007)
28. Rosen, S.: The economics of superstars, The American Economic Review 71 (5), pp. 845–858 (1981)
29. Hall, S., Szymanski, S., Zimbalist, A.S.: Testing Causality Between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer, Journal of Sports Economics 3 (2), pp. 149–168 (2002)
30. Müller, O., Simons, A., Weinmann, M.: Beyond crowd judgments: Data-driven estimation of market value in association football, European Journal of Operational Research 263 (2), pp. 611–624 (2017)
31. Frick, B.: The Football Players´ Labor Market: Empirical evidence from the major european leagues, Scottish J Political Economy 54 (3), pp. 422–446 (2007)
32. Herm, S., Callsen-Bracker, H.-M., Kreis, H.: When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community, Sport Management Review 17 (4), pp. 484–492 (2014)
33. Scelles, N., Helleu, B., Durand, C., Bonnal, L.: Professional Sports Firm Values, Journal of Sports Economics 17 (7), pp. 688–715 (2016)
34. Football Money League, Deloitte (2017)
35. Brand Finance®, (2017)
36. Aral, S., Dellarocas, C., Godes, D.: Introduction to the Special Issue —Social Media and Business Transformation: A Framework for Research, Information Systems Research 24 (1), pp. 3–13. (2013)

37. Singh, S.: Generation Z: Rules to reach the multinational consumer, Insight (2014)
38. Gummerus, J., Liljander, V., Weman, E., Pihlström, M.: Customer engagement in a Facebook brand community, Management Research Review 35 (9), pp. 857–877. (2012)
39. Bühler, A., Balzer, J., Leibbrand, M.: Social Media im Profisport, Nachspielzeit, vol. 2 (2013)
40. Goh, K.-Y., Heng, C.-S., Lin, Z.: Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content, Information Systems Research 24 (1), pp. 88–107. (2013)
41. WELT, https://www.welt.de/sport/fussball/article124066855/Die-groesste-Fussballrevolution-findet-im-Netz-statt.html (Accessed: 08.07.2018)

Local Shopping Platforms – Harnessing Locational Advantages for the Digital Transformation of Local Retail Outlets: A Content Analysis

Sören Bärsch[1], Lars Bollweg[1], Richard Lackes[2], Markus Siepermann[2], Peter Weber[1], and Valerie Wulfhorst[1]

[1] Fachhochschule Südwestfalen, Soest, Germany
{baersch.soeren,bollweg.lars,weber.peter, wulfhorst.valerie}
@fh-swf.de
[2] Technische Universität Dortmund, Dortmund, Germany
{richard.lackes, markus.siepermann}@tu-dortmund.de

Abstract. Competitors and customers put Local Owner-Operated Retail Outlets (LOOROs) under digitalization pressure. Local Shopping Platforms (LSP), acting as intermediaries between LOOROs and their customers, explicitly make use of the locational strength of LOOROs and seem to be a promising vehicle to help LOOROs overcome their manifold digitalization difficulties. In this study, with the help of a structured content analysis of 27 LSPs in Germany, Switzerland, and the U.S., we analyze LSPs as local descendants of e-marketplaces and derive a functionality-based typology. Furthermore, we scrutinize how LSPs harness LOOROs locational advantages. Despite their visible role as inter-organizational service-hubs and a low-level entry option to e-commerce for LOOROs, our results show that LSPs currently unduly focus on location-enabled services that support the online channel, while neglecting the potential of location-based services and the local stores as a Point of Sale (PoS).

Keywords: Local Shopping Platforms, LOOROs, Location-dependent Services, Retail.

1 Introduction

In an otherwise positively developing market environment, Local Owner-Operated Retail Outlets (LOORO) face an intense business and market transformation. The market share of LOOROs has already declined from 26% in 2003 to 17.9% in 2015 in Germany [1]. Several independent studies predict a further decline in revenue for LOOROs in Germany of 30% within the next four years [2], [3] and of about 50% within the next ten years [4]. Big box retail outlets, chain stores and online retailers on

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the one side, as well as changing customer shopping habits and a decline in shopper frequency in the high streets on the other side, are threatening the very existence of LOOROs [5], [6]. Further, studies show that LOOROs, like other small- and medium-sized enterprises (SME), still hesitate to adopt e-commerce channels (e.g. by starting their own online shops) [7], [8] as well as to participate in electronic marketplaces (e.g. selling on eBay or Amazon) [9]. Existing research has explained this phenomenon of reluctant digital transformation with an extensive set of internal (e.g. financial constraints, lack of knowledge, lack of infrastructures, fear of global competition, etc.) and external (e.g. lack of standards, lack of understanding of SME needs, etc.) influence factors acting as adoption barriers [10-13].

Currently, Local Shopping Platforms (LSPs) as intermediaries and inter-organizational service hubs between LOOROs and their customers are spreading in German cities [14]. For instance, Atalanda.com is a local shopping platform provider that offers LSPs in now thirteen German cities. Atalanda enables small and medium sized stationary retailers to sell their products online via local Atalanda marketplaces (e.g. “OnlineCity Wuppertal” or “Einkaufen in Attendorn”).

The advent of LSPs has many ties to the long tradition of e-marketplaces and e-intermediaries with one clear distinction: On large e-marketplaces, like Amazon or eBay, the boundaries between Business-to-Business (B2B) and Business-to-Consumer (B2C) as well as regional or national restrictions blur [15], [16]. With the advent of LSPs, we see a counter-development of platforms that implement location-dependent self-restrictions into their business models. It is either a limitation to the cooperation with retailers from a certain area, the limitation of just doing business with customers from a certain area, or both. LSPs use these location-dependent self-restrictions as their unique selling proposition. They sell the existing offer of merchandise of LOOROs via their platform and try to reach time advantages in delivery and service by using the connected local shops as decentralized storages in direct neighborhood to the households of the local customers [17].

On the one hand, this LSP approach seems to be promising for LOOROs, as it helps them in their digital transformation and to overcome the e-commerce adoption barriers by outsourcing the digitalization challenge to the platforms [18], [9]. The LSPs serve as digital service providers and relieve LOOROs from the burden of building up their own digital infrastructures and hiring specially educated staff [19]. Furthermore, LSPs enable cooperation and shared services among competitors, allowing for synergy effects for the digital transformation of all participating retailers in the city centers. For example, LSPs can spread the development cost of the platform’s infrastructure (IT, logistics) across the connected shops, avoiding the need for individual high investments [20].

On the other hand, local shopping platforms seem to be problematic for LOOROs, as these seem to step into a self-reinforcing spiral of ubiquitous online price competition with significantly declining (online) sales prices when joining the platforms [16]. Furthermore, it remains unclear whether customers will (in the long-term) accept electronic marketplaces that are limited to offers from only local vendors, or whether they will prefer global marketplaces with nearly unlimited offers like eBay and Amazon [21]. Today, the business model of LSPs still lacks proof of concept and is therefore an uncertain bet as a sustainable backbone for the digital transformation of LOOROs.

Against this background, LSPs seem to be a premature approach, but the integration of location advantages and physical infrastructures into their service strategy could lead to competitive advantages for LOOROs. It could help them win back market shares and revenue from the big online and offline competitors and to catch up in the ongoing digital transformation of the retail sector.

However, also reknown pure players, like esp. Amazon, have already started to enter the stationary retail sector (e.g., Amazon Books Stores, Amazon Go Stores, Amazon Now and Whole Foods). These examples of online-to-offline ventures experiment with ambitious retail concepts (e.g., the use of sensors and other technology to identify and automatically charge customers). Therefore, the stakes are high and LOOROs are facing tough competition. As LSPs are a rather new approach, we still do not know enough about their service offers and about how they facilitate the locational advantages of LOOROs in the digital realm. It is not yet to say, if LSPs will actually be able to help LOOROs transform their business models and survive in the highly competitive digital future [19]. Therefore, with regard to the ongoing transformation process of LOOROs and the according role of LSPs, in this paper we aim to answer the following research questions:

RQ1) *What kind of local shopping platform approaches exist?*

RQ2) *Which local services do these local shopping platforms offer in order to utilize the locational advantages of LOOROs?*

The paper is structured as follows: In section 2, we discuss the theoretical background needed for the analysis. We then introduce the methodological foundation of the structural content analysis in section 3. In section 4, we discuss the results of the content analysis. Finally, in section 5, we conclude, point out limitations and discuss future research opportunities.

2 Theoretical Background

2.1 Functionality and Categories of Local Shopping Platforms

Local Shopping Platforms represent e-marketplaces with a local focus. Like traditional e-marketplaces, they serve three main functions: 1) they match (local) buyers and sellers, 2) they enable the exchange of information, and 3) they facilitate transaction and fulfillment services [22], [15].

As with e-marketplaces, also the services provided by local shopping platforms are not standardized and strongly differ [23]. The service landscapes provided range from simple externalized management of the online front-end (e.g., online shop, marketplace) to the outsourced management of complex sales and marketing processes, including pricing, invoicing, data processing, and logistics [23], [24]. According to Peterson et al. (2007) [25], this diverse service landscape enables a typological differentiation (categorization) of LSPs based on the typical e-marketplace functionalities (see Table 1).

Following this approach, we derived five types of local shopping platforms: The first e-marketplace function (match of buyers and sellers) allows for the differentiation of two categories of local shopping platforms, *Store Locator Platforms* and *Product Catalog Platforms*. *Store Locator Platforms* offer only contact and store location information. They do not provide any information on products or any online shop functionalities to enable the core buying process. In addition to this contact and store location information, *Product Catalog Platforms* as the second category of LSPs provide an overview of the products available in the connected stores. As *Store Locator Platforms*, also *Product Catalog Platforms* do not provide any online shop functionalities. Considering the e-marketplace function (exchange of information), we derived an additional platform category named *Product Enquiry Platforms.* These platforms enable customers to request product availability information, while still not providing online shop functionality. Regarding the e-marketplace function (transaction and fulfillment), two more types can be differentiated. *Affiliate Transaction Platforms* allow the purchase of products, but customers complete the transaction process on an external website with the help of an affiliate shop. Finally, *Full Transaction Platforms* offer the full e-marketplace service range, including payment and logistics. Table 1 summarizes the typology.

Table 1: Local Shopping Platform categories with regards to the main e-marketplace functionalities, derived from Peterson et al. (2007)

	Information	**Communication**	**Transaction & Fulfillment**
Platform Categories	Store Locator Platforms	Product Enquiry Platforms	Affiliate Transaction Platforms
	Product Catalog Platforms		Full Transaction Platforms

2.2 Location Theory & Location-Dependent Services

“Location! Location! Location!” has long been a mantra for the stationary retail sector and its service providers. In the pre-e-commerce era, it was widely believed that the choice of a location is the single most important decision for retailers [26]. In fact it has been even argued that poor location may be an insurmountable obstacle for even the best retailers [27]. In the advent of e-commerce, location seemed to have lost its importance [28]. Many pure players built big warehouses outside the cities to enable the efficient implementation of the fulfillment promises they made to their customers [29]. However, with the rising service competition between pure e-commerce players and stationary retailers (Big Box Retail Outlets and LOOROs), including services like e.g. same day delivery or even same hour delivery, the importance of location for the retail sector and the interest in location theory is increasing again [30], [31]. For example, Kim et al. (2017) [32] point out that the impact of distance in the e-commerce age “is not dead”, and in their study they highlight the negative effects of growing transportation costs on online sales prices and customer demand. Furthermore, in their study they show that location-dependent services, like logistics services and especially express delivery services are positively correlated with repurchases and the loyalty of customers. Further, research showed that even in an environment with near zero trade cost, physical distance matters [33] and that especially location-dependent service are essential to attract and retain customers in multi-channel retail environments [34], [35].

To make use of locational advantages, location theory suggests the facilitation of location-dependent services for multi-channel retailers and e-marketplaces. In fact, there are two types of services to look at. First, location-enabled services: these services are possible if the location of the store is close to the households of the customers and thus enables short distance services with low transportation costs [36], [37]. Second, location-based services: These services aim to utilize foot-traffic at popular places like main streets, parks, etc. using location-awareness information systems and devices, like smartphones and wearables [38-41].

Table 2: Differentiation Location-independent and Location-dependent Services

Location-independent Services	**Location-dependent Services**	
Standard Web-Services	Location-enabled Services	Location-based Services
All Services provided on e-marketplaces and not related to location.	Services that are feasible if the location of the retailer is close to the households of the customers.	Services that are feasible if the customers are close to the store location.

3 Analysis

3.1 Methodology

To analyze the different types of existing LSPs and their offered services, we conducted an extensive content analysis [42-44] and followed the procedure suggested by Krippendorff (1980, 2004) [45], [46] and Mayring (2010) [47]. Accordingly, after defining the research scope and questions, in a first step we identified the existing LSPs through an explorative web search. In the second step, we conducted a pre-test, categorizing 30% of the identified LSPs and their services. This pre-test was followed by a revision procedure to improve the categorization and to streamline the coding agenda.

Subsequently, in a fourth step, three individual coders conducted the full content analysis (100%). Finally, in the last step, an expert panel of senior researchers discussed and resolved discrepancies in the coding results (see Figure 1). All mentioned research steps will be discussed in detail in the following.

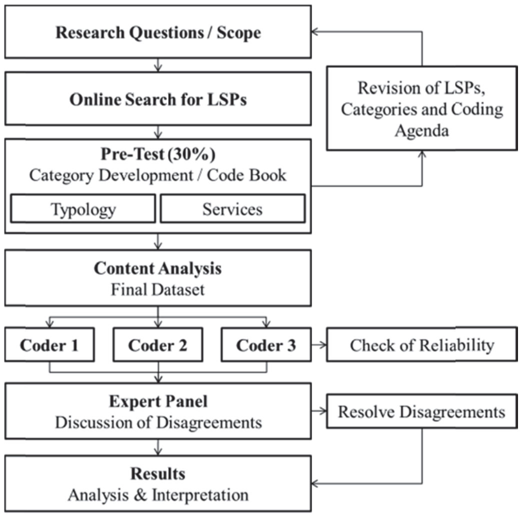


Figure 1: Research Procedure (based on Kripppendorff 2004 and Mayring 2010)

3.2 Sample

For the identification of the LSPs we conducted an online search process in June and July 2016, using the following keyword combinations in English and in German: "Local" + ("E-Marketplace", Shopping Platforms", "Shops Online", "Vendors Online", "Marketplace", "Products Online", "Retail", "Online shop"). The online search process was limited to the first ten result pages on Google per search iteration (first 100 results). To get as unbiased results as possible, the search engine settings where switched to "Do not use private results".

The search process resulted in a first set of 52 candidates for local shopping platforms. A second level selection and screening process to improve the quality of the findings reduced this set to 27 LSPs (selection criterion: local self-restriction to either cooperation with retailers from a certain area, the limitation of just doing business with customers from a certain area, or both) from three countries. 21 platforms were from Germany, four platforms from the U.S. and two platforms from Switzerland (see Table 6). Despite the bilingual keyword combinations and the disabled private results setting, there were distortions in favor of German platforms. The reason for this bias could be default settings of the Google.de search engine.

3.3 Pre-Test, Coding, and Full Content Analysis

After the pre-test (30%) and the revision procedure, the coders together categorized the identified local shopping platforms and their services (100%) in a joint effort to achieve a consistent coding. They screened each platform for 74 possible items: 5

typological items and 69 service items (see Table 5 and 6). Each item was considered with 1 point in the Code Book (Yes / Available = 1, No / Not Available = 0), resulting in two scores: (1) a Typological Score, ranging from 0 to 5, and (2) a Service Score, ranging from 0 to 69 (see Table 3 and 4).

Table 3: Code Book: Platform Typologies

Items / Functions	Definition	Coding Rule
Store location	Shows locations and contact information of local retail stores	Available = 1 Not available = 0
Product Description and Pictures	Shows products and descriptions of products of local retail stores	Available = 1 Not available = 0
Product enquiry	Offers the functionality to write a product enquiry to a local retail store	Available = 1 Not available = 0
Affiliate links	Shows products and prices from local with affiliate link to an external online shop	Available = 1 Not available = 0
Full transaction handling	Offers full transaction handling for local retail stores on the platform	Available = 1 Not available = 0

Table 4: Code Book: Services

Type	Categories	Definition	Items	Coding Rule
Location-Dependent Services	Location-Enabled Services	Services based on the proximity to the customers	16	Available = 1 Not available = 0
	Location-Based Services	Services based on the location of the customer	13	Available = 1 Not available = 0
Location-Independent Services	Information & Recommendation Services	Services offering basic information and / or recommendations	12	Available = 1 Not available = 0
	Communication & Support Services	Services offering a communication channel	14	Available = 1 Not available = 0
	Payment & Billing Services	Services enabling the payment process	8	Available = 1 Not available = 0
	Fulfillment Services	Services for delivery and / or pick-up	6	Available = 1 Not available = 0

3.4 Intercoder Reliability

The verification of the Intercoder Reliability followed the guidelines of Raupp and Vogelsang (2009) [48] and Tinsley & Weiss (1975, 2000) [49], [50]. For this, the Holsti's Coefficient of Reliability $r_{H,}$ and Krippendorffs α were calculated. Concerning the platform typologies, there were 5 items and 27 platforms, so that each coder had to judge 135 different items in total. In sum, complete agreement was achieved ($r_{H,}$=1; see Table 5).

Table 5: Holsti´s Coefficient of the typological items

Coder Pair	**C1 + C2**	**C1+C3**	**C2+C3**
coder's consensus	135	135	135
percentage agreement	1.00	1.00	1.00
r_H	1.00		

Regarding the services, each coder had to examine 27 platforms for 69 services (1,863 items). The Holsti´s reliability coefficient of $r_H = 0.916$ shows an almost complete agreement [51]. The Krippendorff's α coefficient marginally missed the threshold for very good (0.9) but is still good with α = 0.895 [46], [52] (see Table 6).

Table 6: Krippendorffs α and Holsti´s Coefficient of Reliablility r_H

Coder Pair	**C1 + C2**	**C1+C3**	**C2+C3**
Agreement / Coder Pair	1719	1698	1705
Correlation Coefficient / Coder Pair	0.923	0.911	0.915
Holsti´s Coefficient of Reliablility r_H	0.916		
Krippendorff's α	0.895		

3.5 Expert Panel

To overcome and harmonize uncertainties (the coders agreed on 1668 out of 1863 items), an expert panel of four senior researchers with high expertise in the field of e-marketplaces and e-commerce discussed all remaining discrepancies (195 items) and made final decisions. Furthermore, the expert panel reviewed and discussed all platforms and confirmed the coding results.

4 Results

The platform types derived from literature in section 2 could be confirmed in practice. Of the 27 platforms identified, 21 are located in Germany, two in Switzerland, and four in the U.S. Only two platforms represent *Store Locator Platforms* and two others *Product Catalog Platforms*. Six platforms provide the functionalities of *Product Enquiry Platforms*. Four platforms can be considered *Affiliate Transaction Platforms,* and thirteen platforms offer the full range of online shopping transaction and fulfillment services as *Full Transaction Platforms* (see Figure 2). These results clearly indicate that the majority of the platform providers strive to offer complete shopping experiences on their platforms.

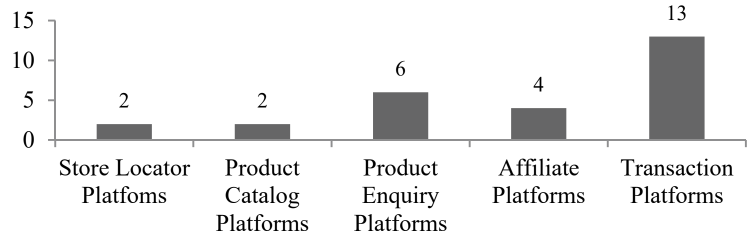


Figure 2: Distribution of types of local shopping platforms in the data sample

Table 7: Identified Location-Dependent Services on LSPs

No.	Location-Enabled Services	#	No.	Location-Based Services	#
	Information Services			**Information Services**	
1.	Map with store locations	16	1.	Location-based product consultation	0
2.	Information about local news	1	2.	Barcode scanner	2
3.	Information about local events	3	3.	Location-based map with store locations	8
4.	Information about product availability (In-Store)	12	4.	Location-based map with closest product location	3
5.	Information about store opening hours	19		**Communication & Support Services**	
6.	Information about store contact data	22	5.	Location-based support	0
	Communication & Support Services		6.	Location-based advertisement	1
7.	Support at home	9	7.	Location-based loyalty program	0
8.	Face to face support	1	8.	Location-based price-draws	0
9.	Local loyalty card	7	9.	Location-based discounts	0
10.	Customer integration (Customer feed-back on store services)	13	10.	In-Store navigation	0
11.	Community integration	10			
	Navigation Services		11.	Outdoor navigation	0
	-		12.	Location-based shopping tour	0
	Payment and Billing Services			**Payment and Billing Services**	
	-		13.	Self-Checkout	0
	Fulfillment Services			**Fulfillment Services**	
12.	Same day delivery	2		-	
13.	Same hour delivery	6			
14.	Click & Return	3			
15.	Click & Collect	0			
16.	Reserve & Collect	6			

Altogether, the 27 platforms provide a range of 67 different services. These services support the different functions of e-marketplaces and can therefore be categorized as follows [15], [22]:

(1) Information & Recommendation Services
(2) Communication & Support Services
(3) Payment & Fulfillment Services

Table 8: Amount of services offered of Location-Dependent and Location-Independent Services on Local Shopping Platforms (July 2016) – Sorted by typology

	Local Shopping Platform			Location-Dependent Services									Location-Independent Services			
				Location-Enabled Services				Location-Based Services								
		Typology	**Total**	**Total LES**	Information	Communication & Support	Fulfillment	**Total LBS**	Information	Communication & Support	Navigation	Payment & Billing	Information & Recommendation	Communication & Support	Payment & Billing	Fulfillment
	No. / category		**69**	**16**	6	5	5	**13**	4	5	3	1	12	14	8	6
	Average no. / LSP		**16**	**4.8**	2.7	0.6	1.5	**0.7**	0.5	0	0.2	0	2	6	2	1
	Share / LSP		**23%**	**30%**	45%	13%	30%	**5%**	12%	0%	6%	0%	20%	41%	28%	15%
1.	Klickando	5	28	**8**	4	0	4	**0**	0	0	0	0	3	9	5	3
2.	Stylerella	5	26	**9**	5	3	1	**2**	1	0	1	0	3	9	2	1
3.	hierbeidir	5	24	**6**	3	0	3	**0**	0	0	0	0	2	8	6	2
4.	Loca Fox	4	24	**6**	4	0	2	**3**	3	0	0	0	6	8	0	1
5.	Atalanda	5	22	**6**	3	0	3	**0**	0	0	0	0	3	7	3	3
6.	farmy.ch	5	22	**5**	1	2	2	**0**	0	0	0	0	3	9	4	1
7.	Kaloka	5	22	**6**	3	0	3	**0**	0	0	0	0	3	7	3	3
8.	buchhandel.de	5	21	**6**	3	1	2	**3**	2	0	1	0	2	4	5	1
9.	Locamo BETA	5	21	**8**	5	1	2	**1**	1	0	0	0	2	6	2	2
10.	Koomio	4	21	**7**	4	0	3	**2**	2	0	0	0	3	8	0	1
11.	Shopcity (Shoplocally)	5	19	**8**	4	4	0	**0**	0	0	0	0	3	5	2	1
12.	findeling	1	17	**4**	4	0	0	**4**	2	0	2	0	0	9	0	0
13.	Digitale City	3	15	**4**	0	1	3	**0**	0	0	0	0	0	8	1	2
14.	locally	3	15	**4**	4	0	0	**0**	0	0	0	0	5	5	0	1
15.	beiuns.kaufen	5	14	**3**	1	0	2	**0**	0	0	0	0	3	5	3	0
16.	Kietzkaufhaus	5	14	**2**	1	0	1	**0**	0	0	0	0	2	5	2	3
17.	localharvestmarket	5	14	**3**	2	0	1	**0**	0	0	0	0	5	3	1	2
18.	Mein Jülich	4	14	**4**	2	1	1	**0**	0	0	0	0	2	8	0	0
19.	Arranja	3	14	**2**	0	1	1	**0**	0	0	0	0	1	8	1	2
20.	RN-Shopping	2	13	**5**	3	1	1	**2**	1	0	1	0	3	3	0	0
21.	Postmates	5	12	**2**	0	0	2	**0**	0	0	0	0	1	5	2	2
22.	snipda	4	12	**5**	4	0	1	**1**	1	0	0	0	2	3	0	1
23.	Lieblingsladen.de	2	12	**4**	3	0	1	**0**	0	0	0	0	3	4	0	1
24.	Marktplatz Bruchköbel	3	11	**7**	4	2	1	**0**	0	0	0	0	2	2	0	0
25.	take-it-lokal.de	3	9	**3**	3	0	0	**0**	0	0	0	0	1	1	4	0
26.	Yategolocal	1	7	**3**	3	0	0	**0**	0	0	0	0	2	2	0	0
27.	Kaufnah BETA	3	3	**0**	0	0	0	**0**	0	0	0	0	0	3	0	0

From the 67 different services, 40 services are location-independent and 29 are location-dependent services. Thirteen services of the latter group are Location-Based Services, while 16 are Location-Enabled Services (see Table 7). Among the location-dependent services, information services are dominating, while among the location-independent services communication services are in the majority.

The extent to which LSPs provide location-dependent services differs quite strong. On average, LSPs offer 23% of all identified services, ranging from a min. of 4% to a max. of 41% of the services. Between 0% and 56% of the identified Location-Enabled

Services (average 30%) and between 0% and 31% of the identified Location-Based Services (average 5%) are provided by the platforms. For comparison, concerning location-independent services, 0%-50% of the information & recommendation services (average 20%), 7%-64% of the communication & support services (average 41%), 0%-75% of the payment & billing services (average 28%), and 0%-50% of the logistics services (average 15%) are provided by the LSPs (see Table 8).

A deeper look on location-enabled services shows that local shopping platforms provide between 0%-83% of the identified information services (average 45%) and 0%-80% of the identified communication & support services (average 13%) as well as of the logistics services (average 30%). As location-based services, only two categories are used: information (0%-75%, average 12%) and navigation services (0%-67%, average 6%).

Logistics services are one of the major advantages that LSPs can offer to customers. Regarding Affiliate and Full Transaction Platforms, which both support the complete shopping process, further interesting findings are that: (1) only seven full transaction platforms offer same day delivery, (2) ten full and two affiliate transaction platforms offer click and collect, (3) five full and one affiliate transaction platform offer click and return, (4) and three full and four affiliate transaction platforms offer reserve and collect.

Regarding the typology, platforms of higher categories tend to use more services than platforms of lower categories. However, this cannot be generalized, as some platforms of category 1 are ranging in the midfield while categories of 5 and 4 can also be found in the lower field.

5 Conclusion

5.1 Discussion

In total, 23 shopping platforms that focus on local retailers and local customers were identified, out of which four are from the U.S. and 23 are from German and / or Switzerland.

Concerning the services provided, there is much room for improvement on local shopping platforms. In average, the examined platforms offer only 16 of the identified 69 services, including only eleven out of 40 location-independent services (27.5%), 4.8 out of 16 location-enabled services (30%), and 0.7 out of 13 location-based services (5%). Thus, although local shopping platforms strive to make use of the locational proximity between shops and customers, they fall short in providing appropriate services. A closer look at location-dependent services reveals that mostly general information regarding shop addresses, contact details, opening hours, and product availability are provided. Most platforms do not offer advanced services that really make use of the position of the customers around or in the local stores connected to the platforms, neglecting the local stores as a PoS. The very few location-based services offered underpin this finding. It almost seems that LSPs are not interested in strengthening the locational position of the connected LOOROs, but only try to benefit from them by focusing on services that foster online sales, like support or payment and billing services.

5.2 Practical implications

For LOOROs: It is questionable if LSPs can help LOOROs with their digital transformation in a sustainable manner. So far, the service offers of LSPs are focusing solely on the online PoS of the platforms itself, while neglecting the digital opportunities of local physical stores. In today's omnichannel retail world with merging sales and communication channels, such a single channel service strategy needs to be considered outdated [21]. Furthermore, big retail chains started to revolutionize the stationary retail sector with innovative store concepts; LSPs seem to be ignoring this upcoming challenge so far [53] (see Table 8).

Despite these critical aspects, LSPs also carry opportunities for LOOROs. The small retailers could make use of the strength of LSPs to get a low barrier entrance to the realm of e-commerce and use the digital environment to learn for their own digital transformation. Accordingly, LOOROs should utilize LSPs as digital service hubs. Furthermore, to improve their role in the relationship with the platforms, LOOROs should demand location-based services from the LSPs to support their physical stores and help to attract more foot-traffic. At the same time, satisfied offline customers have the opportunity to buy 24/7 at their preferred LOOROs via the local platforms. This offers loyal customers an online channel, preventing them from switching to other online shops [35].

For LSPs: With their current business and service strategy, LSPs miss a great opportunity. Instead of solely installing a sales platform that restricts itself and its market to local customers, LSPs should focus more on the locational advantages of LOOROs and foster location-based and enabled services as unique selling proposition [28], [34]. This would strengthen the position of the connected LOOROs in the city centers and help them attract more customers. Such effects are important also for the LSPs, as only strong local retail partners are able to provide a sustainable base for a long-term cooperation, securing the very existence of LSPs. Without support structures that strengthen the local retailers in their ongoing digital transformation, LSPs run the risk of being only a temporal phenomenon.

5.3 Limitations and directions for future research

The results of this study should be interpreted in the context of several key limitations. First, we derived the sample for the analysis from an explorative web search on Google that might have missed additional types of LSPs hidden on the Web. Additionally, the search settings of Google's search algorithm are not fully transparent and therefore the study could suffer e.g. from a country-dependent search bias. Therefore, future studies should consider also other sources of information to identify platforms, like e.g. industry information websites or blogs, and they should work with different language and regional settings on Google. Furthermore, due to the coders' language proficiencies, the analysis was limited to German and English language platforms. Insights about digital platforms for local retailers from other countries would have been of interest, too. Regarding the services offered by LSPs, research is needed on the orchestration of Location-enabled and Location-based Services, and the list of Location Dependent Services in general. Finally, further research is necessary on if and how LSPs and Location Dependent Services can help LOOROs overcome informational disadvantages in comparison to online retailers (e.g. In-Store Analytics via Location-based Services).

References

1. HDE - Retail Federation Germany: Brief Profile. Handelsverband Deutschland, einzelhandel.de/images/presse/Graphiken/HDE_Presentation_eng.pdf/ (Accessed: 31.08.2018)
2. IFH Köln: Stadt. Land. Handel 2020. IFH, http://www.ifhkoeln.de/News-Presse/Fast-jedes-zehnte-Ladengeschaeft-von-Schliessung-bedroht-all (Accessed: 04.08.2016)
3. Heinemann, G.: Online-Handel gräbt stationärem Einzelhandel das Wasser ab –bereits 15 Prozent Anteil in 2013 erwartet. In: eWeb Research Center (eds.). Hochschule Niederrhein, Krefeld, www.hs-niederrhein.de/forschung/eweb-researchcenter/aktuelles/ (Accessed: 31.08.2018)
4. Siemssen, S.: The Oliver Wyman Retail Journal, www.oliverwyman.com/content/dam/oliverwyman/europe/germany/de/insights/publications/2017/Feb_2017/2017_Oliver_Wyman_Retail_Journal_German.pdf (Accessed: 31.08.2018)
5. Bollweg, L., Lackes, R., Siepermann, M., Weber, P.: Mind the Gap! Are local Retailers Misinterpreting Customer Expectations Regarding Digital Services? (Extended Version). IADIS International Journal on WWW/Internet 13(1), 17-29 (2015)
6. Christensen, C. M.: The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. MA: Harvard Business School Press, Boston (1997)
7. Bollweg, L.M., Lackes, R., Siepermann, M., Sutaj, A., Weber, P.: Digitalization of Local Owner Operated Retail Outlets: The Role of the Perception of Competition and Customer Expectations. Pacific Asia Conference on Information Systems (PACIS) (2016)
8. Pantano, E., Viassone, M.: Demand Pull and Technology Push Perspective in Technology-based Innovations for the Points of Sale: The Retailers Evaluation. Journal of Retailing and Consumer Services 21(1), 43–47 (2014)
9. Stockdale, R., Standing, C.: Benefits and Barriers of Electronic Marketplace Participation: an SME Perspective. Journal of Enterprise Information Management 17(4), 301-311 (2004)
10. Vize, R., Coughlan, J., Kennedy, A., Ellis-Chadwick, F.: Technology Readiness in a B2B Online Retail Context: An Examination of Antecedents and Outcomes. Ind. Market. Manage. 42(6), 909–918 (2013)
11. Rahayu, R., Day, J.: Determinant Factors of E-commerce Adoption by SMEs in Developing Country: Evidence from Indonesia. Procd. Soc. Behv. 195, 142–150 (2015)
12. Kabanda, S., Brown, I.: A Structuration Analysis of Small and Medium Enterprise (SME) adoption of E-Commerce: The case of Tanzania. Telemat. Inform. 34(4), 118–132 (2017)

13. Kurnia, S., Choudrie, J., Mahbubur, R.M., Alzougool, B.: E-commerce Technology Adoption: A Malaysian Grocery SME Retail Sector Study. J. Bus. Res. 68(9), 1906–1918 (2015)
14. Bach, C.: Atalanda und Buy Local wollen mehr lokale Online-Marktplätze aufbauen. Location Insider, locationinsider.de/atalanda-und-buy-local-wollen-mehrlokale-online-marktplaetze-aufbauen/ (Accessed: 31.08.2018)
15. Standing, S., Standing, Love, C.: A Review of Research on E-marketplaces 1997–2008. Decis. Support Syst. 49(1), 41-51 (2010)
16. Pan, X., Shankar, V., Ratchford, B.T.: Price competition between pure play versus bricks-and-clicks e-tailers. Advances in Applied Microeconomics 11, 29-61 (2002)
17. Reimann, E.: David gegen Goliath: Wie der lokale Handel gegen Amazon kämpft, http://www.heise.de/newsticker/meldung/David-gegen-Goliath-Wie-der-lokale-Handel-gegen-Amazon-kaempft-2552842.html (Accessed: 04.08.2016)
18. Sandberg, K.W., Håkansson, F.: Barriers to Adapt Ecommerce by Rural Microenterprises in Sweden: A Case Study. International Journal of Knowledge and Research in Management and E-Commerce 4(1), 1-7 (2014)
19. Valenduc G., Vendramin P.: Digitalisation, between disruption and evolution. Transfer Eur Rev Labour Res. 23(2), 121–134 (2017)
20. Huber, B., Sweeney E., Smyth, A.: Purchasing Consortia and Electronic Markets. Electronic Markets 14(4), 284-294 (2004)
21. Brynjolfsson, E., Hu, Y.J., Rahman, M.S.: Competing in the Age of Omnichannel Retailing. MIT Sloan Manage. Rev. 54(4), 23 (2013)
22. Bakos, Y.: The Emerging Role of Electronic Marketplaces on the Internet. Communications of the ACM 41(8), 35-42 (1998)
23. Zheng, W., Wang, X.: An Explorative Study of Industry Influences: On Vertical E-marketplaces' Adoption of E-procurement Auction. Inf. Syst. E-Bus. Manag. 6(4), 321-340 (2008)
24. Wieland, T.: Innenstadtentwicklung und Strukturwandel im Einzelhandel -Ergebnisse einer Angebots-und Nachfragestudie zum Einzelhandel im niedersächsischen Mittelzentrum Holzminden unter besonderer Berücksichtigung der Innenstadt. Universität Göttingen, Göttingen (2011)
25. Petersen, K.J., Ogden, J.A., Carter, P.L.: B2B E-marketplaces: A Typology by Functionality. Int. J. Phys. Distr. Log. 37(1), 4-18 (2007)
26. Craig, C.S., Ghosh, A. McLafferty, S.: Models of the Retail Location Process: A Review. J. Retailing 60(1), 5-36 (1984)
27. Achabal, D.: MULTILOC: A Multiple Store Location Model. J. Retailing 58(2), 5-25 (1982)
28. Iyer, K.N., Germain, R., Frankwick, G.L.: Supply chain B2B E-commerce and Time-based Delivery Performance. Int. J. Distr. Log. 34(8), 645-661 (2004)
29. Weltevreden, J.W.: B2C E-commerce Logistics: The Rise of Collection-and-delivery Points in The Netherlands. Int. J. Distr. Log. 36(8), 638-660 (2008)
30. Allen, J., Piecyk, M., Piotrowska, M., McLeod, F., Cherrett, T., Ghali, K.,Wise, S.: Understanding the Impact of E-commerce on Last-mile Light Goods Vehicle Activity in Urban Areas: The Case of London. Transport Res. D-Tr. E. 61(No. Part B), 325-338 (2018)
31. Lösch, A.: The Economics of Location. Yale University Press, Yale (1954)
32. Kim, T.Y., Dekker, R., Heij, C.: Cross-Border Electronic Commerce: Distance Effects and Express Delivery in European Union Markets, International Journal of Electronic Commerce, 21(2), 184-218 (2017)

33. Blum, B., Goldfarb, A.: Does the internet defy the law of gravity? Journal of International Economics 70(2), 384–405 (2006)
34. Massad, N., Heckman, R., Crowston, K.: Customer satisfaction with electronic service encounters. International Journal of Electronic Commerce, 10(4), 73–104 (2006)
35. Saeed, K.A., Grover, V., Hwang, Y.: The relationship of e-commerce competence to customer value and firm performance: An empirical investigation. Journal of Management Information Systems, 22(1), 223–256 (2005)
36. Jones, M., Mothersbaugh, D., Beatty, S.: The Effects of Locational Convenience on Customer Repurchase Intentions across Service Types. J. Serv. Mark. 17(7), 701-712 (2003)
37. Jensen, C.S.: Research Challenges in Location-enabled M-services. In: Proceedings Third International Conference on Mobile Data Management, pp. 3-7. IEEE Press, New York (2002)
38. Pura, M.: Linking Perceived Value and Loyalty in Location-based Mobile Services. Manag. Serv. Qual. 15(6), 509-538 (2005)
39. Sahai, A., Machiraju, V.: Enabling of the Ubiquitous e-Service Vision on the Internet. E-Service 1(1), 5-19 (2001)
40. Mennecke, B., Strader, T.: Where in the World on the Web Does Location Matter? A Framework for Location Based Services in M-commerce. In: AMCIS 2001 Proceedings, p. 90 (2001)
41. Cabri, G., Leonardi, L., Mamei, M., Zambonelli, F.: Location-dependent Services for Mobile Users. IEEE T. Syst. Man. Cyb. 33(6), 667-681 (2003)
42. Weber, R.P.: Basic Content Analysis (Vol. 49). Sage, Thousand Oaks (1990)
43. Kolbe, R.H., Burnett, M.S.: Content-analysis Research: An Examination of Applications with Directives for Improving Research Reliability and Objectivity. J. Consum. Res. 18(2), 243-250 (1991)
44. Malhotra N.K., Agarwal J. Peterson M.: Methodological Issues in Cross-cultural Marketing Research: A State-of-the-art Review. Int. Market. Rev. 13(5), 7-43 (1996)
45. Krippendorff, K.: Validity in Content Analysis. In: E. Mochmann (Ed.), Computerstrategien für die Kommunikationsanalyse, pp. 69-112. Campus Verlag, Frankfurt am Main (1980)
46. Krippendorff, K.: Content Analysis. An Introduction to Its Methodology (2nd ed.), p. 237. Sage, Thousand Oaks (2004)
47. Mayring, P.: Qualitative Inhaltsanalyse. Handbuch qualitative Forschung in der Psychologie, pp. 601-613. Springer VS, Wiesbaden (2010)
48. Raupp, J., Vogelsang, J.: Medienresonanzanalyse. Eine Einführung in Theorie und Praxis (1. Aufl.). Springer VS, Wiesbaden (2009)
49. Tinsley, H.E., Weiss, D.J.: Interrater Reliability and Agreement of Subjective Judgements. J. Couns. Psychol. 22(4), 358-376 (1975)
50. Tinsley, H.E., Weiss, D.J.: Interrater Reliability and Agreement. In: Tinsley, H.E.A., Brown, S.D. (eds.) Handbook of Applied Multivariate Statistics and Mathematical Modeling, pp. 95-124. Academic Press, San Diego (2000)
51. Holsti, O.R.: Content Analysis for the Social Sciences and Humanities, pp. 138-141. Addison-Wesley, Reading (1969)
52. Neuendorf, K.: The Content Analysis Guidebook. Sage, Thousand Oaks (2002)
53. Perakslis, ED.: Strategies for delivering value from digital technology transformation. Nat Rev Drug Discov. 16(2), 71–72 (2017)

A Socio-Technical Approach to Manage Analytics-as-a-Service – Results of an Action Design Research Project

Christian Dremel[1], Emanuel Stoeckli[1], Jochen Wulf[1], and Walter Brenner[1]

[1] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
{christian.dremel, emanuel.stoeckli, jochen.wulf, walter brenner}@unisg.ch

Abstract. The ability to generate business-relevant information and its use for business process improvements is a key success factor for businesses today. Answering the call for further research on success-relevant practices and instruments for managing business analytics, we report on the results of a three-year action design research project at a global car manufacturer. Drawing on the socio-technical systems theory, we identify seven meta-requirements and specify four principles for the design of an instrument to manage Analytics-as-a-Service (Aaas) portfolios. Our results reinforce the importance of coordinating different socio-technical components in business analytics initiatives and demonstrate how concrete management instruments, such as the proposed portfolio management tool, contribute to socio-technical alignment. For practitioners, the documented design components may provide guidance on how to design and implement similar instruments that support the management of Aaas portfolios.

Keywords: analytics-as-a-service, action design research, big data analytics, business analytics, governance.

1 Introduction

The ability to generate business-relevant information and exploit it to improve business processes is a key success factor for businesses today in facing the digital transformation [1]. Even in industries with commodity products and highly standardized services, business analytics offers companies the opportunity to generate new competitive advantages [2]. The possibilities for use are diverse and range across the individual functional areas of a company (e.g., supply chain, pricing, marketing, product quality, research, and development) with companies such as Amazon or Google to established companies such as Wal-Mart or Novartis, who successfully use analytics to their advantage [2]. Although the trend towards leveraging data to improve decision-making is not entirely new, the technological pervasion through digital technologies such as sensors or personal digital devices (e.g., smartphones and tablets) and the ability to store and analyze data in real-time through innovative technologies led to the emergence of the phenomenon of big data analytics [1]. In order to leverage this disruptive potential and to bundle analytics competencies, larger companies increasingly adopt shared service models and offer Analytics-as-a-Service (AaaS) to

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

business units [3]. However, most traditional companies are unable to generate business value through analytics projects and AaaS offerings, because they fail to address the associated structural, skill-specific, technology-oriented, and business process-oriented challenges stemming from the socio-technical nature of business analytics [4–6]. More precisely, the process of value creation through analytics requires an active and collaborative sense-making process of organizational participants (e.g., analysts and business managers) within the boundaries of the organizational socio-technical work system [7, 8]. Consequently, the management of AaaS not only involves the management of business demands and service requirements but requires an active management of service prototypes and implementations. Such activities are typically handled by a portfolio management function supported by specific management tools or instruments (most particularly a portfolio management tool) [9, 10]. Despite the growing body of practitioner reports [1, 5] as well as of academic literature [11, 12], research on how to exploit analytics to the advantage of companies is still scarce [11, 13] . With our research we address the lack of academic knowledge on portfolio management for AaaS by identifying principles for the design of a portfolio management tool. Accordingly, we pose the following research question: *"How to design a portfolio management tool to support the systematic management of AaaS?"*

We report on the results of a three-year action design research project, which was started at the beginning of 2015. This project aimed at developing a single point of truth for all analytics services to manage analytics services according to the priorizations of the central analytics unit's customers of the organization *CarCo*. At the beginning of our project all services were managed and coordinated with manual documentation on a printed excel spread sheet. Due to the increasing complexities the central analytics unit faced the underlying manual tasks were not feasible anymore and slowed down the day to day work of the central analytics unit. During this project, we used several academic approaches to tackle our phenomenon of interest (e.g., focus group workshops, semi-structured interviews, and observations).

Our work provides two key contributions. First, this work specifies requirements and design propositions for designing an AaaS portfolio management tool and thus addresses the need to further consider "socio-technical and socio-material design considerations for algorithmic decision systems" [14]. Second, this paper contributes rich empirical insights on the (up to this date underexplored) success factors of managing AaaS and responds to calls for further research in this area [11, 14–16].

2 Theoretical Foundations

2.1 A Socio-Technical Systems Perspective on Business Analytics

The socio-technical systems (STS) theory offers a multivariate system perspective on information systems that consists of four interrelated system components and allows its analysis [17, 18]. The social subsystem comprises *structures* and *actors*, whereas the technical subsystem comprises the *technology* and a *task* component [17, 18]. Actors are, amongst others, characterized by employees' capabilities and qualifications as well as a shared culture, structures by project organizations and institutional arrangements, technology by tools and technological platforms, and tasks by the required processes to fulfill work or the delivery of services [17–19].

Following Chen et al. [12] and Holsapple et al. [20] we use business analytics as an umbrella term encompassing the term big data analytics. Business analytics can be understood as a socio-technical phenomenon affecting organizations in multiple ways [20, 21]. Accordingly, to conceptualize analytics from a STS-perspective, (1) the technical system and (2) the social system have to be differentiated. First, the implementation of analytics technology, such as Hadoop cluster and hive (*technology*) is required to provide the technological basis to improve e.g., decision-making processes, and to achieve product, and service innovation (*task*) [1, 22, 23]. Second, in regard to the social system, *actors* require applied skills in the areas and intersections of technology, analytics, and business, others a general knowledge about big data analytics to successfully lead, manage, and implement big data initiatives [24]. Further, an organization needs to transform its organizational *structures* such that they support and establish cross-departmental collaboration [2, 4]. Therefore, we regard analytics as a strategic capability to gain analytical insights from (big) data which equally resides in technological and organizational system components.

In IS research STS theory proved to be useful to elaborate IS induced changes in the organizational context by taking interrelationships between social and technical system components into account [19]. The interactions between the social and technical systems are inherently recursive, that is "users shape the technology structure that shapes their use" [25]. The recursive use of technology in the organizational context results in "enacted structures of technology use" and "while users can and do use technologies as they were designed, they also can and do circumvent inscribed ways of using the technologies" [25]. Thus, knowledge about the alignment of socio-technical components is of critical importance, because information systems are continuously challenged by incremental and punctuated changes to one or more of its system components [19].

This view has general implications for the development of a tool to manage and govern an AaaS portfolio: Before successfully exploiting a technology innovation such as analytics, a multitude of organizational transformations is necessary (e.g., cultural and structural change). Thus, when designing tool to manage and govern an AaaS portfolio the "enacted structures of technology use" [25], which might result from the recursive interaction with analytics services, need to be taken into account. Therefore, to realize the potential advantage of AaaS, organizations need to analyze and understand how they can use and manage data analytics and require associated management instruments. Following the call of Abbasi et al. [15] for further research on data analytics and big data, we use "action design research (ADR) to guide the development and harnessing of big data IT artifacts [in a real-world] organizational [setting]" [15]. In particular, we focus on designing and implementing an instrument to manage the AaaS portfolio at the central analytics unit of a leading car manufacturer using the socio-technical systems theory as reference theory.

2.2 Portfolio Management

In the following, we relate our study to portfolio management, in general, and to project portfolio and service portfolio management, in particular, as these adjacent research fields inform our design of the portfolio management tool for AaaS. The original concept of portfolio management was coined by Markowitz [26]. Portfolio management in its original sense targeted the financial domain to decide on investments portfolios. Over the years, portfolio management was used to manage new product development [27], new service development [28], project management [29–31], and IT service management [32–34]. Extant literature on portfolio management [27, 29, 31] outlines several key objectives of portfolio management: maximization of value (i.e., value maximization of the portfolio against business objectives), balance of the portfolio (i.e., managing the risk of the portfolio), and strategic alignment (i.e., alignment of the portfolio with the overall strategy). Literature on project portfolio management further details these objectives into "(1) defining goals and objectives, i.e., clearly articulating what the portfolio is expected to achieve, (2) understanding, accepting, and making trade-offs, (3) identifying, eliminating, minimizing and diversifying risk, (4) monitoring portfolio performance, i.e., understanding the progress that portfolio is making towards the achievement of the goals and objectives, and (5) establishing confidence in achieving a desired objective" [30]. In targeting the systematic portfolio management of AaaS, we understand service portfolio management "as a dynamic decision-making process that is dedicated to the continuous, strategically aligned revision of service portfolios" [33]. Thereby, several portfolio management models can be distinguished, e.g., financial models, scoring models and checklists, mapping approaches, or mathematical optimization procedures [27]. To achieve this, the portfolio management is widely supported by a dedicated portfolio management tool [28, 30]. Accordingly, the ultimate goal of this research endeavor is the provision of a corresponding portfolio management tool for AaaS.

In summary, the literature provides valuable insights into project and portfolio management related to the definition of projects and of goals, minimization of goals as well as the confidence in these goals [30]. However, due to the socio-technical nature of analytics the realization of the potential business value from analytics, socio-technical configurations, amongst others related to a data-oriented practices, structures, technologies, and analytical processes need to be taken into account when pursuing analytics projects and developing a portfolio management tool for AaaS [7, 8]. For this reason, previous approaches are not applicable in this context without adapting or extending these approaches.

3 Research Approach

We conduct action design research (ADR) with the goal to obtain relevant results by means of a rigorous yet pragmatic approach due to its suitability for addressing practice-inspired research problems [35, 36]. By drawing on the existing body of knowledge and by guiding the artifact building and evaluation phases, ADR supports the development of prescriptive design knowledge. ADR aims at generating solutions that apply not only to a problem instance but to a class of problems [36]. We embedded the 3-year ADR project in a collaborative practice-oriented research initiative [37] and followed Sein et al.'s [36] methodological guidelines (e.g., practice-inspired research, theory-ingrained artifact, and reciprocal shaping) (see Figure 1).

At the outset, we founded a collaborative research initiative with *CarCo*, a leading innovative original equipment manufacturer in the automotive industry who was on the edge of launching an AaaS initiative. *CarCo* is a multinational car manufacturer who targets customers in the premium segment and is known for their technological advancement and produces around 2 million cars a year making a profit of around €4 billion and employing around 90000 people world-wide. Particularly, we collaborated with several members of the central analytics unit at *CarCo* (among others the Head of Data Strategy and Analytics), who represent the targeted users of our artifact - in our case the portfolio management tool. We discussed and evaluated preliminary results in the build phase of the ADR project with practitioner experts. Thereafter, the prototype was iteratively built, evaluated, and refined with the targeted users, i.e., the Head of Data Strategy and Analytics, and stakeholders, i.e., the Head of Analytics Services.

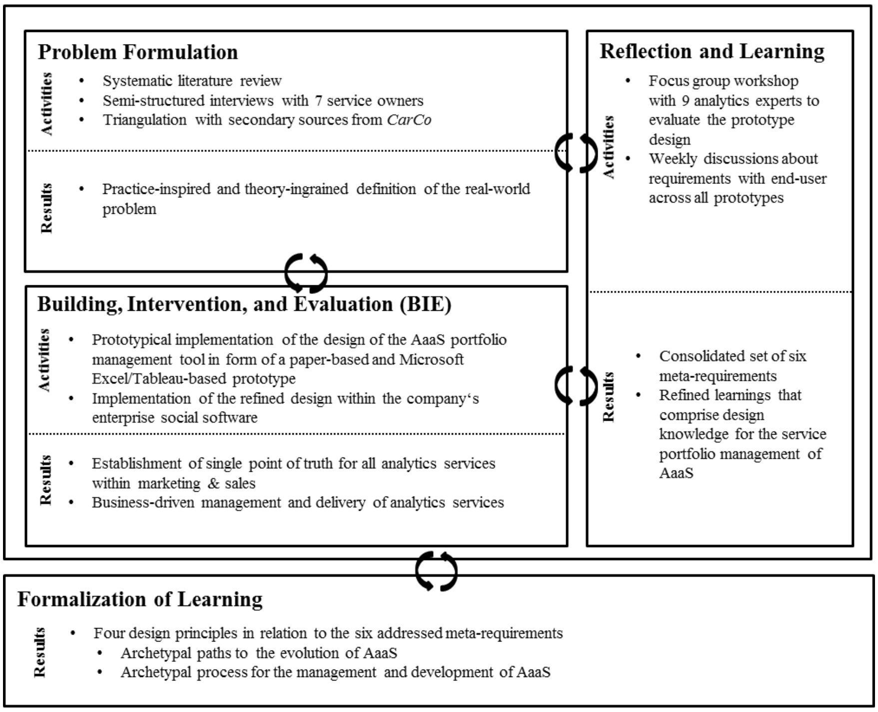


Figure 1. Overview of ADR stages with Key Activities and Results [36]

Following the ADR methodology and particularly Sein et al.'s [36] principle of 'guided emergence', meta-requirements (MR) emerged from the reflection on and learning from the problem formulation phase. In the subsequent building, intervention and evaluation phases, a solution design was iteratively refined in a three-cycle design process. The concurrent reflection on the design in our real-life application context resulted in learnings that comprise generalized design knowledge for the service portfolio management of AaaS to address not only a problem instance but also a class of problems [38]. Thus, we elicit MRs that apply to the problem class that the ADR project aims to address (i.e. AaaS portfolio management). To do so, we follow the two key principles of Sein et al. [36]: (1) practice-inspired research and (2) theory-ingrained artifact. To elicit MRs, we use triangulation and rely on rich data from both (1) a systematic literature review and (2) data obtained from a focus group workshop with domain experts, and multiple bilateral interviews with targeted end-users (i.e., Head of Data Analytics & Strategy). Finally, these learnings were formalized as generic design principles (DP) addressing the identified MRs to generate prescriptive design knowledge.

Problem Formulation. First, to identify potentially relevant related work, we conducted a literature search following the guidelines of Webster and Watson [39]. We were interested in identifying prior business analytics portfolio management approaches, following Sein et al.'s [36] principle of drawing on prior theory ("Principle 2: Theory-Ingrained Artifact"). We broadly searched the literature base for articles addressing the management of business analytics and analytics services. As keywords, we used the terms: "business intelligence" and "analytics" in combination with the more general terms "service" and "portfolio management" to query the databases "EBSCO Host Business Source Complete", "ProQuest", and "AIS library" resulting in 22 hits for "analytics" and 9 hits for "business intelligence".

We, initially, screened all articles in detail to sort out irrelevant articles. However, we were not able to identify reusable concepts for our phenomenon of interest (i.e., the management of an AaaS portfolio) in the current body of literature. As a consequence, we turned to more general theories on business analytics, its antecedents and contextual factors and portfolio management in general. In particular, we turned to IT service management [32–34], project portfolio management [29–31], and product portfolio management [27]. Since business analytics represents a socio-technical phenomenon that affects organizations in multiple dimensions (e.g., in regard to culture, organizational structures, work processes, capabilities, or, ultimately, business models) [1, 21, 23, 40] and to follow the principle of a "theory-ingrained artifact" [36], our research is informed by the socio-technical systems (STS) theory. In doing so, we are able to increase rigor within the problem formulation phase of our action design research project. Drawing on these foundations we conducted interviews with seven service owners, which were responsible for the management and development of seven analytics services within the AaaS initiative. We collected meta-requirements in the four STS-dimensions and used Atlas.ti for analyzing and coding interview notes. Third, secondary data that had been provided either by the interviewees or the manager in charge of the AaaS initiative further augmented our empirical data. As a result of data triangulation, we derived MRs and refined them in a focus group workshop with nine experts on business analytics to pursue the process of ex-ante evaluation of our prescriptive design knowledge [41]. These experts were a Performance and Improvement Manager as well as a Product Manager Business Intelligence from the agriculture industry, a Head of Digital Market Intelligence from market research industry, a Product Manager Business Intelligence & Data Science as well as a Chief

Executive Officer of an analytics advisory from retail industry, a Head of Digital Marketing from lightning industry, a Product Manager Data Management as well as a Product Manager Business Intelligence from the fast moving consumer goods industry, and a Professor for Digital Business. To do so, a paper-based prototype consisting of 15 data fields each corresponding to our initial set of MRs was explained to the participant of focus group workshop. In a next step, the participants gave formal feedback through a handout as well as informal feedback in case further explanations were needed.

Building, Intervention, and Evaluation (BIE). We performed an ex-post naturalistic evaluation drawing on interviews and observations as evaluation methods [41, 42]. To do so, the identified MRs were addressed in a first cycle by means of a prototypical paper-based implementation. We applied and refined at *CarCo's* analytics unit based on multiple user feedback cycles with the Head of Data Analytics & Strategy. In a second cycle a prototypical portfolio management tool was designed using Excel's functionality to not only store data but also to create user interfaces and manipulate data through its programming language Visual Basic for Applications and Tableau as visualization software to make the artifact tangible for the end-users (managers of the central analytics unit and the product management of the AaaS initiative) allowing the evaluation of its practical use [41]. Thus, after this initial implementation of the management tool, it was evaluated and validated with the Head of Analytics Services during a two-hour meeting and with the Program Manager for the AaaS initiative [36] during an additional two-hour meeting. This resulted in the consolidation of our initial set of 15 meta-requirements to 6 MRs for the following reasons. First, our initial set of MRs did consider the service instance on a fine granular level, which increased the effort to maintain the service portfolio. Second, due to the feedback of the focus group workshop rather redundant fields were eliminated (e.g., "business problem" as it was perceived as too detailed and too redundant to the existing field "business value"). Acknowledging the consolidated set of 6 MRs, the tool was finally in a third cycle integrated into *CarCo's* enterprise social software (ESS) as part of the AaaS Atlassian Confluence Space of the central analytics unit. During this implementation, we constantly evaluated in face-to-face meetings our artifact with the Head of Data Analytics & Strategy, the Head of Analytics Services, and, primarily, the product owner of the analytics-as-a-service initiative by ensuring the practical use and foremost feasibility to maintain and sustain the artifact after our action design research project.

Formalization of Learning. To provide prescriptive design knowledge we derived design principles through three additional interviews representing market participants (i.e., Head of Artificial Intelligence and Machine Learning at *VehicleCo*, Head of Data Analytics at Marketing and Sales at *VehicleCo*, Managing Consultant Development of Analytics Services at *PremCo*), as well as one interview with the program manager of the AaaS initiative drawing on our previous research steps. *VehicleCo* and *PremCo* can be seen as the most important competitors to *CarCo* targeting the same customer segment and producing cars in a similar engineering quality. As a result of analyzing this interview data and reflecting the project results, we were able to identify four design principles that reflect core characteristics of the proposed instrument for AaaS portfolio management. We acknowledged during these interviews that not only *CarCo* but also

VehicleCo and *PremCo* follow an archetypal path for the initial development of AaaS and sourcing decision. Moreover, in a similar fashion depending on their sophistication in regard to technology (i.e., technological readiness), analytical expertise and analytical capabilities (i.e., analytical readiness), and domain knowledge three archetypal paths dominated how analytics services are managed through their classes of services (i.e., descriptive, predictive, and prescriptive analytics services).

4 Results

4.1 Problem Design and Problem Solution

In the following, we elaborate on the six meta-requirements, which are based on our theoretical foundation (i.e., STS theory) and result from the reflection on the problem formulation phase (see Table 1):

Table 1. Overview of Identified Meta-Requirements

ID	Description	STS Domain
MR1	The service description must include a definition of the analytics service in order to achieve a common understanding of each service's functionality and outcome.	Task
MR2	The business value of each analytics service needs to be described.	Structure
MR3	The artifact should allow maintaining the feasibility and costs of the analytics service.	Structure
MR4	The artifact should allow to maintain the roll-out status and plan depending on the real-world context and the level of granularity.	Task
MR5	Service owners from both the business side as well as the analytics experts should be identified for each analytics service instance.	Actor
MR6	The description of each analytics service should include the analytical type (i.e., descriptive, predictive, and prescriptive) and the analytical method.	Technology

MR1. Following the need for a structured service description for service portfolio management [33] in general and in particular for analytics projects [7], we formulate MR1 based on our empirical findings as follows: "The service description must include a definition of the analytics service in order to achieve a common understanding of each service's functionality and outcome." Addressing MR1, the data fields service name, service definition, and comment are included to ensure a unified understanding of each service. The service definition describes the service functions in a business-oriented manner and thus refers to the task dimension of the socio-technical system.

MR2. To fulfill one of the main purposes of portfolio management (i.e., strategic alignment) [27, 30, 31, 33] – in our case to strategically align analytics services with the overall strategy - and grounded in our empirical insights, we introduce MR2: "The business value of each analytics service needs to be described". In addition, knowledge about the inherent value of analytics projects plays a crucial role to communicate service innovation possibilities [7, 8]. Addressing MR2, each generic service of the service portfolio is mapped to its corresponding value driver (i.e., business value) and business area (e.g., marketing & sales) enabling a prioritization of services. A strict value-orientation allows to govern analytics services in accordance with business strategy and thus affects the structure dimension of the socio-technical system.

MR3. To allocate a company's resources, align the portfolio with an organization's capacity, and to avoid losing sight of the costs [30, 31, 34] and drawing on our empirical evidences MR3 is theorized as "The artifact should allow maintaining the feasibility and costs of the analytics service." In particular, as analytics projects are evolutionary in nature the continuous control of costs is key for the management of AaaS. Addressing MR3, an initial estimate of the cost, as well as the complexity of the implementation of the analytics service, is provided by the management instrument. This supports communication and work flows in the governance dimension of the social-technical system. The ability to track and manage feasibility of the analytics service is particularly important for the end-user of this artifact as the provision of analytics services is charged to of internal business customers.

MR4. Due to the diverse customers of the AaaS initiatives (i.e., different target groups as part of CarCo or external, such as dealers and importers of CarCo) different granularity levels of services and their current status need to be taking into account [32–34]. In particular, CarCo's dealers and importers are distributed across the globe and do not form part of CarCo's organization. Therefore, they also might use services from external partner more easily. In result, the marketing of the analytics services based on the conducted implementation activities and the rollout status is highly important. Therefore, MR4 is phrased as "the artifact should allow to maintain the roll-out status and plan depending on the real-world context and the level of granularity." Answering to MR4, the rollout status of each service instance is documented within the service inventory table. The rollout status describes conducted implementation activities and thus relates to the task dimension in the socio-technical system.

MR5. To align business and IT, to assure a clear allocation of responsibilities, and to achieve confidence in the service [30, 32, 33] as well as due to the cross-disciplinary nature of analytics [4] MR5 is theorized as "Service owners from both the business side as well as the analytics experts should be identified for each analytics service instance." In response to MR5, each service instance has one service owner from the business department at the demand-side (i.e., a business owner) and one service owner stemming from the analytics experts at the provider-side (i.e., analytics owner). The specification of service owners relates to the actor dimension in the socio-technical system.

MR6. Allowing the prioritization and management of services depending on their current class of service [32, 33] MR6 is phrased as follows: "The description of each analytics service should include the analytical type (i.e., descriptive, predictive, and prescriptive) and the analytical method." Guided by MR6, each service is mapped to the analytics type (i.e., descriptive, predictive, and prescriptive) and the analytics model (e.g., decision-tree). These descriptions of the technological underpinnings relate to the technology dimension in the socio-technical system.

Instantiation of the Problem Solution. The artifact itself comprises a service portfolio management tool aiming at representing the single point of truth for all analytics services within marketing and sales. For instance, the AaaS offering includes a service to analyze depending on CarCo's data the demographic distribution of potential leads to increase the lead conversion as well as to potentially achieve up-selling – in the case of CarCo to transfer a customer from one car model to a superior car model. As such it follows a similar goal as project portfolio management. Namely, a centralized view of all projects – in our case analytics services - in an organization. Following Cooper et al. [27] the service portfolio management tool is a portfolio management tool to map current analytics services according to their strategic prioritization and to index analytics services according to their current status. As a result, analytics services are maintained by the analytics unit in the service portfolio management tool and a customer-orientated delivery of analytics services to internal business customers can be achieved.

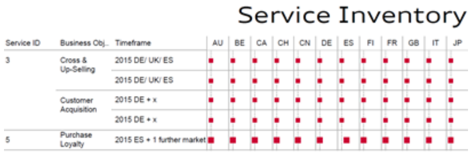


Figure 2. Extract of the Prototype (only non-confidential information visible)

The derived meta-requirements (see Table 1) are met through specific data fields, which are needed to describe an analytical service on a detailed level. Being initially implemented with a Tableau dashboard and an Excel spreadsheet using VBA for data entry (see Figure 2) the artifact was finally implemented within the space of the central analytics unit's enterprise social software (see Figure 3). In doing, so the final artifact was accessible for all relevant personnel, namely the product owners of each service. They were tasked with maintaining the key information of the artifact and were updated to any changes made regarding their services due the clear assignment through the field "responsible" leveraging the notification system of Atlassian Confluence.

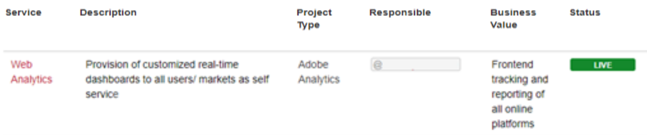


Figure 3. Extract Service Overview of the Live Version (only non-critical information visible)

4.2 Formalization of Learning

Contributing to the scientific body of knowledge on service portfolio management tools for AaaS, the design knowledge obtained within the ADR project is formalized in general design principles. These design principles represent learnings and generalized knowledge of the solution that was built within the course of the ADR project [36]. Table 2 provides an overview of the derived design principles and relates them to the addressed meta-requirements. They are further discussed in the following.

Table 2. Overview of Identified Design Principles [36]

ID	Design Principle	Addressed MRs	STS Domain
DP1	Maintain a taxonomy that guides the specification of business objectives of an AaaS offering in the portfolio management instrument.	MR1 MR2	Structure, Task
DP2	The demand-side business owner of an AaaS instance must be documented within the portfolio management instrument. Further, this business owner must be accepted by the supply-side.	MR3 MR5	Structure, Actors
DP3	For portfolio management to identify implementation status and to guarantee service quality, the portfolio management instrument must allow to assess how the implementation of each AaaS complies with a standardized archetypical development process.	MR3 MR4 MR6	Actor, Task
DP4	To ensure the successful implementation of AaaS, portfolio management instrument must allow the consideration of the trinity of technological readiness, analytical readiness, and business acumen.	MR3 MR5	Actor, Technology

DP1. The first design principle ensures sufficient visibility and budget for the development of AaaS offerings. Since in early project phases it is often non-trivial to define feasible and realistic business objectives, the maintaining of a taxonomy of AaaS business objectives and examples supports the specification of business objectives related to a planned or realized AaaS offering:

"You need the business case, you need the customer. Just to take data, to analyze it and to show the results is not enough. This approach often failed in the past because the business units just have no interest." Product Owner Analytics Services #1, CarCo, 2016

DP2. The second design principle targets a close collaboration of both the analytics experts (possibly part of business departments or IT department) and the requesting business units. A demand-side owner of the service instance should be specified for portfolio management. Further, this business owner should be accepted by the supply-side as a respected and qualified counterpart:

"The projects are always strong when a strong business unit is on board and that is something, we learn little by little in the projects. That means those topics that we started in a data-driven manner, and about which we had information only based on our available data, mostly did not make it far. But if we have someone from the demand-side, who is excited about the topic, and a business owner, who is familiar with the business problem, and combine this with analytical expertise, then it works." Head of Data Analytics & Strategy, CarCo, 2016

DP3. Relating to the third design principle**,** an archetypical process for the development of AaaS supports portfolio management in specifying implementation stage and in guaranteeing service quality: Triggered by a crisp and clear project idea in mind, the phase of exploration starts. Within this phase, the strategic relevance of the corresponding idea is assessed together with the internal feasibility (i.e., it is assessed if it can be implemented as AaaS internally). In this regard, an initial assessment of the technological readiness (i.e., data sourcing, access, integration, and delivery), the available or potential analytical expertise and analytical capabilities (i.e., analytical readiness) is conducted. If the project idea is assessed as strategically relevant and feasible, then, an internal implementation is pursued as far as possible. Non-strategic topics, however, are outsourced due to their minor business relevance. This principle highlights the critical importance of how well the business department understands the business value they want to achieve (i.e., business acumen and understanding).

"We particularly select the cases with a high strategic impact and a high complexity in regard to the analytical method necessary. In doing so, we encourage and challenge our employees to make sure we develop the best people" Head of Artificial Intelligence & Machine Learning, VehicleCo, 2017

DP4. When pursuing AaaS, an organization has to develop corresponding knowledge and capabilities. This includes: (1) data sourcing, access, integration, and delivery as well as connected data sources and continuous data streams (i.e., technological readiness), (2) analytical expertise and analytical capabilities (i.e., analytical readiness), and (3) domain knowledge to understand what is hidden behind the company's data (i.e., business understanding and acumen). Portfolio management must assess these capabilities since they determine implementation costs and the evolution paths of business analytics services. For instance, as *CarCo*, *VehicleCo*, and *PremCo* matured with respect to these three dimensions, three archetypical paths became evident in the empirical data. These pathways characterize how an analytics service can be further developed depending on an organization's maturity on technological readiness, analytical readiness, business acumen and understanding (due to page limitations these pathways are not illustrated in detail). For instance, to acknowledge these pathways the portfolio management tool references to and documents the underlying technological infrastructure (e.g., data source systems and streams) to consider the technological readiness. Further, the responsible data scientists and developers are documented to ensure the consideration of the analytical readiness. Finally, the business uses cases, their responsible employee, as well as a product owner from the analytics unit to consider aspects of business understanding form part of the portfolio management tool's data fields.

5 Conclusion

Although business analytics receives considerable attention in practitioner and academic literature, little research is providing empirical evidence on value realization and utilization so far [16]. Building upon the holistic enterprise perspective of the socio-technical systems theory, we report on the results of a 3-year ADR project, in the course of which we designed and implemented a database that supports the management of an AaaS portfolio. In this article, we illustrate our problem design and solution and present the accumulated knowledge, which results from the formalization of learnings. There are three key limitations in the light of which our research results have to be interpreted. First, our research was conducted in form of a collaborative research initiative. Since ADR projects may suffer from a bias related to the subjectivity of the involved researchers, we included researchers from outside of the project in our team and involved multiple researchers in the coding of our qualitative data. Second, since this

ADR project was exclusively conducted at a single company, our results may suffer from limited generalizability. To ensure that our proposed solution applies to a class of problems (and not only to the specific problem instance at CarCo), we involved external experts in multiple phases of the research project (e.g., focus group workshop with 9 analytics experts and three additional interviews with *VehicleCo* and *PremCo*). Third, guided by the extant body of knowledge as well as the project-based approach of *CarCo* to the management of AaaS our results bear similarities to the existing body of knowledge (e.g., requirement of a detailed service description). With our meta-requirements and design principles for the design of a service portfolio management tool for AaaS, we contribute a nascent design theory [43] to the emerging literature on AaaS [44, 45]. In particular, we address a lack of research on instruments and practices for the management of business analytics to realize the business value [11, 13]. Given the important role of portfolio management for the development of successful service offerings [9, 10], we develop a prescriptive instrument for the management of AaaS portfolios informed by STS theory. Ignoring the socio-technical complexity of implementing analytics services is a main reason for project failures [5, 6], our holistic approach covers all components of a socio-technical system. Particularly, the meta-requirements focus on the individual STS components (see Table 2). The design principles target the alignment and interrelationship of individual STS components (see Table 3). For example, conjunctively nominating demand-side business owners (DP2) leads to an alignment of governance instruments (i.e., the structure component) and involved actors. Similarly, assessing the technological readiness leads to an alignment of the actor's technology capabilities (i.e., the actor component) with the applied analytical instruments (i.e., the technology component). The results of our ADR projects reinforce the importance of coordinating the different socio-technical components in business analytics initiatives. Our contribution to practice lies in providing actionable guidance on how to implement a database that supports the management of an AaaS portfolio.

References

1. Davenport, T.: Big data @ work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press (2014).
2. Davenport, T.H., Harris, J.G.: Competing on analytics: The new science of winning. Harvard Business Press (2007).
3. Dremel, C., Herterich, M., Wulf, J., Waizmann, J.-C., Brenner, W.: How AUDI AG Established Big Data Analytics in Its Digital Transformation. MIS Q. Exec. 16, 81–100 (2017).
4. Sharma, R., Mithas, S., Kankanhalli, A.: Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. Eur. J. Inf. Syst. 23, 433–441 (2014).
5. Demirkan, H., Dal, B.: The Data Economy: Why do so many analytics projects fail?, http://analytics-magazine.org/the-data-economy-why-do-so-many-analytics-projects-fail/.
6. Singh, P.: 10 Reasons Why Do Big Data And Analytics Projects Fail?, https://analyticsindiamag.com/10-reasons-big-data-analytics-projects-fail/.

7. Troilo, G., De Luca, L.M., Guenzi, P.: Linking Data-Rich Environments with Service Innovation in Incumbent Firms: A Conceptual Framework and Research Propositions. J. Prod. Innov. Manag. 34, 617–639 (2017).
8. Dremel, C., Herterich, M., Wulf, J., Brocke, J.V.: Actualizing Big Data Analytics Affordances: A Revelatory Case Study. Inf. Manage. (forthcoming).
9. Ajjan, H., Kumar, R.L., Subramaniam, C.: Information technology portfolio management implementation: a case study. J. Enterp. Inf. Manag. 29, 841–859 (2016).
10. Jeffery, M., Leliveld, I.: Best practices in IT portfolio management. MIT Sloan Manag. Rev. 45, 41 (2004).
11. Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., Zhao, J.L.: Transformational Issues of Big Data and Analytics in Networked Business. MIS Q. 40, 807–818 (2016).
12. Chen, H., Chiang, R.H.L., Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Q. 36, 1165–1188 (2012).
13. Tambe, P.: Big Data Investment, Skills, and Firm Value. Manag. Sci. 60, 1452–1469 (2014).
14. Galliers, R.D., Newell, S., Shanks, G., Topi, H.: Datification and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making. J. Strateg. Inf. Syst. 26, 185–190 (2017).
15. Abbasi, A., Sarker, S., Chiang, R.H.: Big Data Research in Information Systems: Toward an Inclusive Research Agenda. J. Assoc. Inf. Syst. 17, 3 (2016).
16. Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M., Feldberg, F.: Debating big data: A literature review on realizing value from big data. J. Strateg. Inf. Syst. 26, 191–209 (2017).
17. Bostrom, R.P., Heinen, J.S.: MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. MIS Q. 1, 17–32 (1977).
18. Leavitt, H.J.: Applied organizational change in industry: Structural, technological and humanistic approaches. In: Handbook of organizations. pp. 1144–1170. Rand McNally & Co., Chicago (1965).
19. Lyytinen, K., Newman, M.: Explaining information systems change: a punctuated socio-technical change model. Eur. J. Inf. Syst. 17, 589–613 (2008).
20. Holsapple, C., Lee-Post, A., Pakath, R.: A unified foundation for business analytics. Decis. Support Syst. 64, 130–141 (2014).
21. Boyd, D., Crawford, K.: Critical Questions for Big Data. Inf. Commun. Soc. 15, 662–679 (2012).
22. Lehrer, C., Wieneke, A., Vom Brocke, J., Jung, R., Seidel, S.: How Big Data Analytics Enables Service Innovation. J. Strateg. Inf. Syst. (2018).
23. Loebbecke, C., Picot, A.: Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. J. Strateg. Inf. Syst. 24, 149–157 (2015).
24. Chatfield, A., Reddick, C., Al-Zubaidi, W.: Capability Challenges in Transforming Government through Open and Big Data: Tales of Two Cities. In: Proceedings of the 36th International Conference on Information Systems, Fort Worth, USA (2015).
25. Orlikowski, W.J.: Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. Organ. Sci. 11, 404–428 (2000).
26. Markowitz, H.: Portfolio selection. J. Finance. 7, 77–91 (1952).
27. Cooper, R.G., Edgett, S.J., Kleinschmidt, E.J.: Portfolio management for new products. Perseus Pub, Cambridge, MA (2001).
28. Aas, T.H., Breunig, K.J., Hydle, K.M.: Exploring New Service Portfolio Management. Int. J. Innov. Manag. 21, 1 (2017).
29. Martinsuo, M.: Project portfolio management in practice and in context. Int. J. Proj. Manag. 31, 794–803 (2013).

30. Reyck, B.D., Grushka-Cockayne, Y., Lockett, M., Calderini, S.R., Moura, M., Sloper, A.: The impact of project portfolio management on information technology projects. Int. J. Proj. Manag. 23, 524–537 (2005).
31. Stettina, C.J., Hörz, J.: Agile portfolio management: An empirical perspective on the practice in use. Int. J. Proj. Manag. 33, 140–152 (2015).
32. Comerio, M., Batini, C., Castelli, M., Grega, S., Rossetti, M., Viscusi, G.: Service portfolio management: A repository-based framework. J. Syst. Softw. 104, 112–125 (2015).
33. Kohlborn, T., Fielt, E., Korthaus, A., Rosemann, M.: Towards a service portfolio management framework. In: Proceedings of the 20th Australasian Conference on Information Systems. pp. 861–870. Monash University (Online), Monash University, Melbourne, Victoria (2009).
34. Peppard, J.: Managing IT as a Portfolio of Services. Eur. Manag. J. 21, 467–483 (2003).
35. Iivari, J.: A paradigmatic analysis of information systems as a design science. Scand. J. Inf. Syst. 19, 5 (2007).
36. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action Design Research. MIS Q. 35, 37–56 (2011).
37. Mathiassen, L.: Collaborative practice research. Inf. Technol. People. 15, 321–345 (2002).
38. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Q. 28, 75–105 (2004).
39. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. Manag. Inf. Syst. Q. 26, xiii–xxiii (2002).
40. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. 52, 21 (2011).
41. Sonnenberg, C., vom Brocke, J.: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Peffers, K., Rothenberger, M., and Kuechler, B. (eds.) Proceedings of the 7th Design Science Research in Information Systems, Las Vegas, USA (2012).
42. Venable, J., Pries-Heje, J., Baskerville, R.: A Comprehensive Framework for Evaluation in Design Science Research. In: Peffers, K., Rothenberger, M., and Kuechler, B. (eds.) Proceedings of the 7th Design Science Research in Information Systems, Las Vegas, USA (2012).
43. Markus, M.L., Majchrzak, A., Les Gasser: A Design Theory for Systems That Support Emergent Knowledge Processes. MIS Q. 26, 179–212 (2002).
44. Delen, D., Demirkan, H.: Data, information and analytics as services. Decis. Support Syst. 55, 359–363 (2013).
45. Naous, D., Schwarz, J., Legner, C.: Analytics As A Service: Cloud Computing and the Transformation of Business Analytics Business Models and Ecosystems. In: Proceedings of the 25th European Conference on Information Systems, Guimarães, Portugal (2017).

Characterizing Approaches to Digital Transformation: Development of a Taxonomy of Digital Units

Christoph Fuchs[1], Philipp Barthel[1], Ina Herberg[1], Matthias Berger[1], and Thomas Hess[1]

[1]LMU Munich, Institute for Information Systems and New Media, Munich, Germany
{fuchs,barthel,matthias.berger,thess}@bwl.lmu.de
{ina.herberg}@campus.lmu.de

Abstract. Confronted with the imperatives of an increasingly digital world, organizations are challenged to maintain the exploitation of existing revenue sources while simultaneously exploring novel paths for a digital future. One option to manage this organizational ambidexterity and to foster innovation activities within the companies is to implement digital units. However, although the introduction of digital units has increasingly become common practice for organizations, a high uncertainty about the nature of such units remains. Therefore, we develop a taxonomy to characterize digital units by building on pertinent literature in the fields of digital transformation, organizational ambidexterity, and organizational design. In combination with employing a qualitative-empirical research approach, we contribute to existing literature by offering an initial characterization of digital units and a first empirical application of our taxonomy. We also provide descriptive findings on digital units in practice and offer insights for companies that consider to implement such specific units.

Keywords: Digital Transformation, Digital Units, Organizational Ambidexterity, Taxonomy Development, Qualitative Case Studies.

1 Introduction

Emerging advancements in information technology (IT) continuously facilitate the development of new digital products, services, and business models [1], [2]. Digital technologies impact markets by increasing their transparency, thus resulting in lowered markets' entry barriers and enhanced competition [3]. Consequently, in such a *digital world*, so far successful business models are threatened to be disrupted by new market entrants, forcing incumbent companies to rethink how business value is created [3], [4].

To remain competitive in such a volatile business environment, companies are urged to continuously renew themselves. Besides other aspects, a constant organizational progress includes the adoption of new technological achievements for the development of digital innovations [5-7]. Thereby, a digital innovation can be understood "as the creation of (and consequent change in) market offerings, business processes, or models that result from the use of digital technology" [8, p. 224]. To enable innovation activities and explore paths for a digital future (e.g., trends such as agile methods [7]),

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

companies increasingly create specific units, so-called *digital units* [9]. Digital units enable companies to realize organizational ambidexterity. While core organizations focus on operating the traditional business and leveraging existing revenue sources (i.e., *exploitation*), digital units focus on innovation and exploration activities in their search for new revenue sources (i.e., *exploration*). Consequently, digital units represent one vital option for managing companies' digital transformation [5], [9].

Existing scientific literature sparsely touched the specific topic of digital units, but addressed essential aspects of the research field. For instance, literature investigates the influence of digital technologies on the nature of innovations and the innovation process itself [2], [8]. Digital units can also be part of digital transformation strategies (DTSs) that aim to govern the digital transformation of companies from a strategic perspective [10], [11]. Additionally, research addresses the question whether companies need to establish the specific role of a Chief Digital Officer (CDO) [12], [13]. However, current literature lacks research on how to engage and anchor the digital transformation in organizations – besides the introduction of a single management position such as the CDO. With digital units representing one option to structurally embed the digital transformation within companies, the emerging organizational designs appear to be an interesting and current research topic from a scientific as well as practical perspective.

A related research field where the necessity for ambidextrous approaches is already addressed is literature on the restructuration of companies' IT functions. Such *bimodal IT organizations* typically encompass a *traditional IT function*, which is responsible for the operation of companies' core IT systems, and a *digital* or a*gile IT function*, which focuses on digital innovation activities [14], [15]. Whereas Horlach et al. [14] clarify the concept of bimodal IT, Jöhnk et al. [15] develop a taxonomy for agile IT design options. Although this research is akin to the approach of implementing digital units, the topic of bimodal IT and agile IT setups has a distinct perspective focusing on the organizational IT function. However, such an IT-focused lens may miss out on some of the diverse forms of digital units realized in practice (e.g., innovation labs, incubators) [9], [16]. In contrast to this focus on the IT function and the related IT ambidexterity, we aim to employ a broader lens that captures the digital innovation activities in light of the entire company and considers approaches to realize organizational ambidexterity.

Although scientific literature discusses important aspects of the fields of digital transformation and bimodal IT, empirical research focusing on the specific issue of digital units and their establishment in companies is still sparse. Therefore, we aim to contribute by developing an initial understanding of what digital units are and how they can be characterized. It appears fruitful to start with a descriptive approach and build initial knowledge in a novel research field [15]. Consequently, we can provide a theoretical basis for future research that examines digital units through an explanatory lens. We derive as a guiding research question: *How can digital units be characterized?*

To answer this question, we aim to develop a taxonomy of digital units. This taxonomy shall reflect pertinent organizational design options of digital units as they are realized in practice. The taxonomy's frame is initially derived from literature by deducing relevant categories. These are then revised and broken down into dimensions and corresponding characteristics based on observations of real-world digital units. To examine these units, we employ a qualitative approach by compiling five case studies.

2 Underlying Research Foundations

2.1 Impact of the Digital Transformation

Specific environments require specific *organizational designs*, for example in times of rapid technological change [17]. The implications of an ongoing digital transformation for markets and companies (e.g., products, business models, operations) reach beyond resource digitization and process automation by means of IT [1], [5]. Consequently, the digital transformation resembles an *IT-enabled organizational business transformation* [18], [19] that can be interpreted as a specific environment which necessitates a unique organizational design such as the implementation of digital units.

It may also be of interest to examine whether the implementation of a digital unit in an organization is part of a holistic DTS [11]. This refers to the origin of digital units and represents a specification whether the units' implementation results from a top-down plan or emerges bottom-up from individual initiatives [20].

2.2 Digital Transformation and Organizational Ambidexterity

Organizational ambidexterity describes the ability of companies to be successful in competing in mature markets while at the same time exploring new markets with flexibility and experimentation [21], [22]. These challenges of ambidexterity can also be transferred to the digital transformation context. Companies need to, for instance, invest in the development of innovative digital business models (i.e., exploration), while at the same time operating their (mostly) non-digital core business (i.e., exploitation) [23]. A prominent approach to achieve organizational ambidexterity is the alteration of the companies' *organizational design* by means of creating separate units for exploration and exploitation [24]. These units are not only structurally separated, but also differ with regard to, for instance, competencies and processes [22], [24]. Exploratory units, such as digital units, typically show higher degrees of freedom and autonomy from the core organization and allow companies to escape the inertia of existing business operations and organizational structures [24], [25]. Accordingly, the creation of separate units in the context of the digital transformation can be observed in practice and sparks organizational units such as innovation labs and incubators [12].

2.3 Understanding the Organizational Design of Digital Units

To examine how digital units can be characterized, we consider design parameters that determine organizational units. As a result, we assume that organizational units require 1) an *objective*, 2) *resources*, and 3) a *structure* to exist within companies [26], [27]. The objective describes the purpose for which the units are created and represents the central determinant for the overall design. This is especially true for units that are designed for a specific purpose, such as in the case of digital units [27]. Subsequently, units require resources to be capable of operating. Besides the financial resources, organizational units also need access to human resources, since the personnel's skills, knowledge and experience are a critical determinant of the effectiveness of units [26], [28]. Additionally, there is also a need for structure, both in the internal (inside the organization) and external (outside the company) function of organizational units [28].

2.4 Practical Approaches to Digital Units

Practitioners have made initial efforts to characterize digital units. For instance, the consulting company etventure offers a differentiation into four types: innovation labs, company builders, incubators, and digital units [16]. The main differences are whether the innovation activities happen within (e.g., innovation lab) or outside of the core organization (e.g., incubator). A similar categorization is offered by Ramus and Velten [29] referring to innovation labs, company builders, accelerators, and incubators. There also exist more fine-grained categorizations that use several layers to define various types of digital units. For instance, Sindemann and von Buttlar [9] separate between digital innovation units (e.g., innovation labs, accelerators) and special forms (e.g., customer co-creation labs, digital venture capitalists). Although these contributions by practitioners are relevant, the proposed characterizations are neither consistent nor theoretically sound. Therefore, we explore the characterization of digital units on a general level without specifying types and thus refer to the general term of *digital units*.

2.5 Initial Grasp of Digital Units

To summarize, the specific organizational design of exploratory units, here digital units, represents one approach to address the digital transformation of companies. Based on the discussed research foundations, we derive an initial understanding of such digital units that guides our study and our qualitative-empirical research approach. Consequently, we comprehend digital units as organizational units that 1) primarily focus on *innovation* and *exploration activities* to pave the companies' paths for a digital future, 2) have *secured access* to financial as well as personnel resources, and 3) possess high degrees of freedom within the respective companies to operate *autonomously*.

3 Research Method

3.1 Taxonomy Development Approach

We develop a taxonomy for the characterization of digital units, because this research approach enables us to offer initial theoretical and empirical insights in an area that is currently predominantly shaped by consulting articles (e.g., [9], [29]).

In our approach, we followed the taxonomy development method proposed by Nickerson et al. [30]. In a first step, it is essential to determine the meta-characteristic of the taxonomy which should "reflect the purpose of the taxonomy" [30, p. 343]. Consequently, reflecting our research question, we defined as our meta-characteristic: *Characteristics of digital units*. Additionally, we assumed that *dimensions* exist that contain a spectrum of coherent characteristics. With the taxonomy development approach being of iterative nature, it is crucial to define *ending conditions* – distinguished into *objective* and *subjective* – that determine a stop to the iteration cycles. We defined the following objective ending conditions: a) No new characteristics were added in the last iteration, b) no characteristics were merged or split in the last iteration, and c) every characteristic is unique and not repeated. In accordance with prior research [15], we did not follow the objective ending condition that characteristics need to be mutually exclusive. Otherwise, combinations of characteristics would need to be included by means of individual characteristics what contradicts the prerequisite of taxonomies to be comprehensive and parsimonious [30]. Therefore, we allowed for combinations of characteristics within one dimension to portray the nature of digital units. Regarding the subjective ending conditions we built on the suggested conditions of Nickerson et al. [30] (i.e., taxonomy is concise, robust, comprehensive, extendible, and explanatory) and formulated a holistic subjective ending condition. This end was represented by the point where all observed real-world cases could be classified in the taxonomy and where we did not have to implement further alteration in the taxonomy.

Since little data about the phenomenon is available, we followed the recommendation of Nickerson et al. [30] and began with a *conceptual-to-empirical* iteration. Here, we reviewed general literature on organizational ambidexterity (e.g., [22], [23], [25]) and organizational design (e.g., [17], [28]), as well as specific articles on digital transformation (e.g., [11]), and practical contributions on digital units (e.g., [9], [29]) to define a frame for the taxonomy. We searched the literature for elements commonly considered as relevant for characterizing organizational designs in general (e.g., objective of organizational units) and digital units in specific (e.g., possible origin from DTSs). Next, we clustered these elements according to their *thematic fit*, resulting in superordinate *categories*. These categories were then extended and refined in the subsequent *empirical-to-conceptual* iteration by identifying coherent dimensions and corresponding characteristics of digital units.

3.2 Case Study Research Approach

For this second, *empirical-to-conceptual*, iteration, we selected an explorative, qualitative-empirical research approach by conducting five case studies on organizations that implemented digital units. Case studies are especially suitable for recent phenomena that should be studied within their real-life context such as the phenomenon of digital units [31]. To increase the study's robustness and enable cross-case analysis, a multiple-case design was selected [31]. We followed a theoretical replication logic to generate contrasting results by choosing diverse cases, thereby enhancing the external validity of the study [31]. Our case selection process was primarily guided by our initial understanding of digital units (see 2.5). We employed a criterion sampling logic and searched for companies that implemented digital units as part of their organizational digital transformation. Additionally, we looked for organizations where the operations of the digital units reached beyond their conceptualization [31]. We identified the manufacturing industry as especially relevant. Respective companies are usually slow to react to the implications of the digital transformation since the mostly non-digital business models are not yet threatened [32].

Our sample consists of five cases that stem from the manufacturing industry, but have distinct areas of operations. Although all companies have a business-to-business focus, they differ in size, experience with and approaches to the digital transformation and their digital units. An overview is shown in Table 1.

Our data collection took place January 2018 – March 2018. We conducted between one and three semi-structured interviews per case. For each case, we aimed to capture the insights of a position responsible for the digital unit (e.g., CDO, head of digital unit) and (if possible) one permanent employee (PE) of the unit. Interviews ranged 45 – 60 minutes and were conducted face-to-face or via telephone. The interviews followed a guide with open-ended questions comprising sections about the digital transformation of the companies, the implementation history of the digital units, and the units' work processes. All interviews were recorded and transcribed verbatim [33]. We triangulated our primary data by secondary data (e.g., firm websites, press releases) to increase the validity of the study [33]. ATLAS.ti was utilized to pool and analyze our gathered data.

Considering the data analysis approach, we followed a two-step approach. First, two researchers independently evaluated the interviews and jointly condensed the collected data to cases. Secondly, we analyzed the cases in terms of the taxonomy development. Hereby, we followed a descriptive coding approach with codes deduced from the categories (i.e., result of the conceptual-to-empirical iteration). During this second step, codes for the derivation of dimensions and characteristics were inductively added. This step was performed by three researchers independently, whereas significant differences in code application were discussed collectively and resolved consensually [33].

Finally, the resulting taxonomy was employed to classify the digital units of the five case studies. To consistently match statements from our data to the dimensions and characteristics, we defined coding guidelines which we refined in the coding process. All classifications were again carried out by three researchers independently, whereas varying assessments of the units' characteristics were discussed and resolved collectively. To complement this empirical-to-conceptual iteration, we reached out to our interview partners and presented them the results of our taxonomy development and the characterization of their digital unit. The subsequent discussions only sparked minor alterations in terms of, for instance, wording of dimensions or characteristics.

Table 1. Overview of the sample

	Case A	**Case B**	**Case C**	**Case D**	**Case E**
Focus within Manufacturing Industry	Steel Processing	Machines for Food Industry	Specialty Chemicals	Investment Holding	Tools
Revenue (2017)	≈ 41 bn. €	≈ 3.7 bn. €	≈ 14 bn. €	≈ 3.6 bn. €	≈ 1 bn. €
Employees (2017)	≈ 160,000	≈ 15,000	≈ 36,500	≈ 17,700	≈ 3,000
Foundation Year DU*	2016	2016	2017	2016	2017
Employees DU*	3	≈ 8	30	8	≈ 20
Number/Dates of Interviews	2/01.2018	1/02.2018	2/01.2018	2/03.2018	3/01.2018
Interviewees by Position *DU is short for Digital Unit	· Head of DU · Project Member (Core Org.)	· Head of DU	· CEO of DU · PE	· Project Lead · PE	· Head of DU · Project Lead · PE

4 First Iteration Cycle: Foundation for the Taxonomy

Reflecting the conceptual-to-empirical iteration cycle, we derived an initial set of categories as foundation for our taxonomy. In this section, we discuss the identified categories against their literature background and in light of digital units.

Objective & Scope. As we are concerned with exploratory units, we set a special focus on the digital units' scope in the innovation process [34]. In addition, the orientation of the units' innovation activities – whether they are directed to external markets or to internal organizational improvements – needs to be specified. Owing to the specificity of the digital transformation, organizations with primarily externally directed digital units may also face the decision whether the focus of innovation activities lies on exploring solutions for existing business areas, discovering new markets, or a combination of these two alternatives [4], [5]. Besides innovation activities, there can be further objectives such as the renewal and expansion of existing core competences and the exploration of current management trends [7].

Staffing & Collaboration. Staffing of digital units and respective projects is expected to be critical. Deploying employees from the core organization may provide company-specific knowledge and an intra-organizational social network, which can be valuable for integrative activities. Delegated employees of the core organization can also acquire novel digital capabilities and transfer them back to the organization. However, often, not all knowledge and capabilities required for exploratory activities can be found within the company. Therefore, external collaborations may be important and a combination of internal and external personnel appears fruitful [24], [25], [35].

Funding. As digital units' projects affect companies as a whole, a logical choice is to offer central funding from the core organization. Digital units would then be cost centers like other support functions, whereas the units' budget is regularly reviewed and the units' heads have to justify spending [25]. A decentralized alternative would be to make different departments provide a share of their budget to the digital units. A third option is that if the units' projects include the commercialization of innovations and thus generate external revenues, it is possible to set them up as profit centers [24].

Governance & Structure. The concept of organizational ambidexterity builds on the assumption that exploratory and exploitative units should be separated to be successful [25], [35]. Both objectives can thus be pursued simultaneously in appropriate settings spanning distinct management concepts, organizational structures and even organizational cultures [23]. However, the approaches to achieve this separation and the degrees of freedom the exploratory units have can vary. Therefore, diverse modes for structurally embedding digital units are possible, ranging from their integration in the core organization to spinning them out as separate legal entities [24], [35]. Exploratory units can also be dissolved or re-integrated after a certain time-span [23]. Besides an organizational separation of units, the physical separation of exploratory units (e.g., distinct location from the core organization) has been considered as a relevant measure to shield innovation activities from the constraints and routines of the core organizations [36]. Additionally, the degrees of freedom of digital units are majorly related to the decision-making power of the digital units' management teams (e.g., power over project topics, applied management concepts, or resource use) [25].

Origins. DTSs are supposed to coordinate, prioritize, implement, and govern a company's digital transformation. Next to the changes in value creation, use of technology, and financial aspects, one important element of DTSs is the aspect of structural changes within the organizations. These changes can, for instance, comprise the implementation of a digital unit. Therefore, we consider companies' DTSs as prevalent origins of digital units, reflecting top-down decisions. Digital units can also emerge from bottom-up initiatives that are later incorporated into a DTS [10], [11].

5 Second Iteration Cycle: Refinement and Application of the Taxonomy

In the following, we present the results of the empirical-to-conceptual iteration cycle and thus the developed taxonomy (see Table 2) as well as its application on our cases. This includes the individual description of all five cases as well as a cross-case analysis.

Table 2. Taxonomy of digital unit's characterization

Category	Dimension	Characteristic				
Objective & Scope	Main Objectives	Digital Innovation	Cultural Change	Development of Digital Expertise		
	Innovation Orientation	Purely Internal	Primarily Internal	Balanced	Primarily External	Purely External
	Market Focus of Innovation	Existing Business Areas	Novel Business Areas			
	Scope of Innovation Process	Idea Generation	Idea Selection	Innovation Development	Innovation Implemen-tation	Innovation Commer-cialization
Staffing & Collabo-ration	Staffing (Project)	Digital Unit Employees	Core Organization Employees	External Partners		
	Importance of Ext. Partners	None	Low	Medium	High	
Funding	Funding (Project)	Central Funding	Business Department	Sponsorship Model	Internal Cost Allocation	External Revenue
Governance & Structure	Embedding	Integrated	Separate Department	Separate Legal Entity	Virtual	
	Permanent	Yes	No			
	Location	Onsite	Offsite			
	Degrees of Freedom	Very Low	Relatively Low	Balanced	Relatively High	Very High
Origins	Origin from DTS	Yes	No			
	Formation	Top-Down	Bottom-Up			

Case A. The digital unit of case A focuses on the development and implementation of digital, digitally enriched and non-digital innovations to explore novel business areas. The unit also aims at accelerating the core organization's cultural change and at building up and transferring digital expertise to the rest of the company. However, no specific activities are undertaken to achieve these additional goals. The projects of the digital unit are initialized and funded by the core organization's higher-level managers. Due to this sponsorship model, the digital unit's degrees of freedom are limited, for instance, the unit does not decide on its project portfolio. Instead, only projects that are sponsored by the core organization's managers are undertaken. Every six months, the sponsors decide on the follow-up financing of projects based on their market chances. The projects of the digital unit are carried out by employees of the core organization who work part-time in the unit's projects. Thereby, the project teams are guided by the digital unit's managing team. The work on the innovation projects takes place in the premises of the core organization and in the co-working space of the digital unit, which is located offsite. The digital unit of case A is not a separate entity, but is a virtual organization[1] located within the company's internal training unit. Consequently, every six months, the continuation of the digital unit's program depends on the availability of core organization managers who are willing to fund projects. The idea to establish a digital unit was launched bottom-up by two members of the core organization and the decision to implement the digital unit was not taken by the executive management.

Case B. The innovation activities of case B's digital unit are limited to the early stages of the innovation development process (i.e., idea generation, selection and prototype development) and focus on existing business fields and the customer side. Additional objectives of the unit include the acceleration of the core organization's cultural change and the acquisition as well as transfer of digital expertise to the core organization. While the former is pursued by conducting workshops, the latter is not pursued by any dedicated activities. Projects to develop innovations are carried out by PEs of the digital unit and employees of the core organization, which are involved part-time in the unit's projects. A central committee consisting of members of the core organization decides on the undertaken projects in the unit. The digital unit is not provided with a budget to carry out projects independently, but instead only approved projects (by the committee) are funded centrally. From an organizational point of view, the digital unit of case B is embedded in the core organization's research and development department, but is physically located offsite in an entrepreneurship center together with other companies. The decision to set up a digital unit was taken at a strategy conference by the organization's board of directors.

Case C. The digital unit of firm C engages in a variety of innovation activities. Among these are the development and implementation of novel digital products and services for new markets, the ideation and testing of corresponding new business models, the leveraging of existing business fields by enriching the respective products and services digitally, the acquisition of pertinent start-ups, as well as the investment in digital pioneers. With its work, especially with projects that take place in cooperation with employees of the core organization, the digital unit aims to transfer digital expertise to the core organization. The acceleration of the core organization's cultural change is excluded from the digital unit's objectives, since it is considered as a task of the digital unit's Chief Executive Officer (CEO) in his role as the core organization's CDO. The unit, which is established as a Limited Liability Company (LLC), is provided with an overall budget from central funds, but also has a profit responsibility and thus generates revenue that it can reinvest in its own projects. The digital unit manages its

[1] In this case, it is an intra-organizational virtual organization that arises due to the task-oriented networking of employees of different business departments.

own project portfolio. Additionally, the unit has co-working spaces in the vicinity of the corporate headquarters and near a branch in China. The digital unit was founded to help the realization of the previously devised DTS of company C. Establishing the digital unit in its existing form was a top-down decision by the executive management.

Case D. Company D's digital unit aims to improve internal business operations, advance existing business fields, and explore new markets. Its primary focus thereby is on the development and implementation of new business models. By involving employees of the core organization in the unit's work, the digital unit aims to accelerate the cultural change of and transfer digital expertise to the core organization. In the course of an innovation's development, the responsibility for it is gradually transferred from the unit to the respective business area in the core organization. Consequently, the projects are staffed with members of the digital unit (e.g., project leads), employees of the core organization (part-time), and external partners. The latter serve the purpose of relieving the digital unit's employees of their work and thus enable several parallel projects within the unit. Additionally, the external partners contribute required skills that are not available in the company. Projects are selected by managers of the core organization and digital unit members jointly. With company D being an investment holding, its digital unit is funded like a business department whereas the core organization provides the financial resources for each project. The work space of the digital unit is designed as a co-working space and is located offsite. The digital unit of company D is an LLC. The board of directors initiated the digital unit at the same time as the organization's DTS has been developed and the unit was integrated into the DTS.

Case E. Company E's digital unit aims to develop novel digital products and services, enrich existing offerings digitally, and establish new digital business models that are related to the organization's core business. Besides this customer centric focus, further activities target internal process improvements. Affecting the company's cultural change and fostering know-how transfer are defined as additional objectives and the digital unit's employees pursue these goals with dedicated activities such as conducting workshops and trainings. The unit engages in the entire innovation process from idea generation to go-to-market strategies and thus is also responsible for the operation of new digital products and services. As a separate unit with responsibilities for its developed products and services, the digital unit operates on a revenue-based model where the core organization provides fixed budgets and treats the digital unit as a profit center. The projects are mostly staffed with PEs of the digital unit and only a small portion of core organization employees. External support is mostly included for missing skills in the project teams. The digital unit of case E is located at the company's headquarter. The unit acts with relatively high degrees of freedom since it determines the products to develop as well as the corresponding development plans autonomously. The decision to establish the digital unit was part of the organization's DTS.

Cross-case Analysis. By comparing our cases, it can be seen that the primary *objective* of all five digital units is the development of digital innovations. Here, the cases vary accordingly whether these innovations are solely externally directed at the targeted markets (i.e., cases A-C) or include the improvement of internal business operations (i.e., cases D and E). The market focus of the case companies differs. Whereas company A solely aims at innovations in novel business fields and organizations B and E focus on their existing business fields, the companies C and D emphasize innovation activities in both business areas. Almost all observed companies aim to employ their digital units as vehicle to accelerate the organizational cultural change and to foster digital expertise in the core organization. Only company C excludes the cultural change as topic for the digital unit, however, this goal is addressed in the core organization by the CDO. While companies B and E use workshops to foster organizational change, all companies rely on interdisciplinary projects with PEs of the digital unit and members of the core organization to transfer knowledge and digital expertise. In the cases A, B, and D the respective digital units gradually hand over the responsibility of the digital innovations to the core organization and do not participate in the actual operation of the products and services. Only in the case of company E, the digital unit takes responsibility in the commercialization of innovations by developing respective strategies and operating products and services. Owing to its diverse activities, the digital unit of company C also covers the commercialization step of the innovation process by means of acquiring and integrating start-ups and the respective products in the portfolio of the organization.

For *staffing and collaboration* of the digital units' projects, all case companies rely on interdisciplinary project teams consisting of PEs of the digital unit, the core organization, and external partners. Owing to the sponsorship model in case A, the projects are performed by core organization members but guided by the digital unit.

This *funding* model of company A is unique, since most digital units' projects are centrally funded by their companies (i.e., cases B, C, E). Owing to the nature of company D as an investment holding, the implementation of the digital unit as a business department enables agency within the diverse structures of the company.

Considering the *governance and structures* of the digital units, all digital units with the exception of the ones of company A and E are exclusively located outside the organizations' headquarters and are designed as co-working spaces. Owing to six-months funding cycles of case A's digital unit, it has to be interpreted as the only non-permanent option. Besides the virtual organization of the digital unit at company A, we found diverse structural embedding of digital units across the cases. Whereas the digital unit of organization B is integrated in the research and development department, the units of companies C and D represent separate legal entities manifested in their status as LLCs. The digital unit at company E is also a separate department, however still legally integrated in the core organization. Albeit these diverse structural embedding options, we cannot derive a direct link between the structures and the degrees of freedom that the units possess. For instance, the integrated digital unit at company B does not possess significantly less autonomy than the two LLCs. Since both digital units at company C and E operate as profit centers and generate own revenues by means of the commercialization of their innovations, these units enjoy higher degrees of freedom, for instance by independently deciding on their project portfolios.

In four of five cases the digital units are part of a *DTS* and were installed on basis of a *top-down decision*. Only the digital unit of company A is not part of the company's DTS. Although other companies (i.e., case D) integrated the already existing digital units in their DTS, the digital unit of case A still lacks a board-level decision which also reflects its bottom-up nature. An overview of our findings can be seen in Table 3.

Table 3. Overview of the cross-case analysis

Dimension	Case A	Case B	Case C	Case D	Case E
Main Objectives	Innovation Cultural Change Digital Expertise	Innovation Cultural Change Digital Expertise	Innovation Digital Expertise	Innovation Cultural Change Digital Expertise	Innovation Cultural Change Digital Expertise
Innovation Orientation	Purely Ext.	Purely Ext.	Purely Ext.	Primarily Ext.	Primarily Ext.
Market Focus of Innovation	Novel Business Areas	Existing Business Areas	Novel / Existing Business Areas	Novel / Existing Business Areas	Existing Business Areas
Scope of Innovation Process	Idea Generation – Innovation Implementation	Idea Generation – Innovation Development	Idea Generation – Innovation Commerc.	Idea Generation – Innovation Implementation	Idea Generation – Innovation Commerc.
Staffing (Project)	Dig. Unit Emp. Core Org. Emp.	Dig. Unit Emp. Core Org. Emp.	Dig. Unit Emp. Core Org. Emp. Ext. Partners	Dig. Unit Emp. Core Org. Emp. Ext. Partners	Dig. Unit Emp. Core Org. Emp. Ext. Partners
Importance of Ext. Partners	Low	None	Medium	High	Medium
Funding (Project)	Sponsorship Model	Central Funding	Central Funding Ext. Revenue	Business Depart.	Central Funding Ext. Revenue
Embedding	Virtual	Integrated	Sep. Legal Ent.	Sep. Legal Ent.	Sep. Depart.
Permanent	No	Yes	Yes	Yes	Yes
Location	Onsite / Offsite	Offsite	Offsite	Offsite	Onsite
Degrees of Freedom	Balanced	Balanced	Relatively High	Balanced	Relatively High
Origin from DTS	No	Yes	Yes	No (Now Integrated)	Yes
Formation	Bottom-Up	Top-Down	Top-Down	Top-Down	Top-Down

6 Implications, Limitations, and Future Research

We contribute to literature by providing a theoretically sound and empirically derived taxonomy for the characterization of digital units. Referring to our research question – *How can digital units be characterized?* – we initially characterized digital units as organizational units that 1) primarily focus on *innovation* and *exploration activities*, 2) have *secured access* to financial and personnel resources, and 3) possess high degrees of freedom within the respective companies to operate *autonomously* (see 2.5).

Referring to characteristic 1), we found that digital units can also have the purpose to support the cultural change of the core organizations and to build up and transfer digital expertise. However, their main focus is on developing and implementing digital innovations for customer markets. These activities can aim to expand existing business fields or explore novel business areas. In light of characteristic 2), all but one observed digital unit (i.e., case A) have secured financial funding for their projects and a stable embedding in the structures of the core organizations. We found that the digital units' projects are typically interdisciplinary staffed with members from the unit and from the core organization. The importance of external partners varied widely across the cases, allowing no general statement. Considering characteristic 3), we found empirical evidence that puts the originally derived high degrees of freedom into perspective. Although most observed digital units were located offsite, only two units (i.e., cases C and E) can, for instance, autonomously direct their project portfolios, whereas the projects of the remaining digital units are either entirely determined by (i.e., cases A and B) or jointly with the core organizations (i.e., case C). This can be attributed to the observation that almost all digital units result from top-down decision and are integrated in organization-wide DTSs. Therefore, it is appropriate to state that digital units operate autonomously *within* the organizational scope that is given by the companies.

Consequently, we define digital units as organizational units with the goal to foster the organizational digital transformation by 1) performing innovation and exploration activities in existing and/or novel business areas. Digital units 2) possess a structured organizational embedding and a secured access to financial resources, 3) conduct projects together with the core organization, 4) act autonomously within the given scope, and 5) are part of company-wide strategies addressing the digital transformation.

Considering practical implications, on basis of our taxonomy, managers can observe the possible design options for establishing digital units within their organizations. Our data shows that there are diverse setups enacted in practice and we could not identify an overall best practice to establish a digital unit. Instead, responsible managers need to purposefully evaluate their companies' specific requirements and implement digital units accordingly. Thereby, it is important to note that our derived dimensions and thus the respective characteristics are not entirely independent from each other. For instance, an interdisciplinary staffing with members of the digital unit, core organization's members and external partners can foster the development and transfer of digital expertise. Additionally, the selected funding option of the digital unit's projects has implications for its degrees of freedom. A sponsorship model, for instance, limits the opportunity for the digital unit to manage its own project portfolio. However, the implementation of a digital unit as a profit center seems only expedient if the unit can, for instance, decide on undertaken projects and has the opportunity to commercialize its innovations. Also, central funding can lead to conflicts between executive management's rather short-term goals to the digital unit's mid- to long-run profitability.

Our findings are not without limitations. Although we followed the guidelines for the development of a taxonomy rigorously [30] and refined our insights on basis of qualitative data, we cannot guarantee that we captured all categories, dimensions, and characteristics to define digital units. In addition, our empirical findings are derived from large companies located in the manufacturing industry with a business-to-business focus, thus limiting the overall generalizability of our results. Therefore, capturing small and medium-sized companies as well as organizations from different industries and with diverse business focuses may spark further insights for the taxonomy. However, as demanded by literature [30], our taxonomy is extendible in regard to further categories, dimensions, and characteristics. Additionally, although we refer to possible connections of various dimensions and respective characteristics, we do not thoroughly investigate the relationships between certain dimensions and characteristics. Therefore, this appears as a fruitful approach for future empirical research on digital units' characterization. Also, in contrast to existing consulting articles (e.g., [9], [29]), we do not derive archetypes of digital units that follow a specific categorization in terms of our developed taxonomy. The development of such a typology requires additional empirical data applied on our taxonomy. Such archetypes of digital units may also potentially be used as blueprints for establishing digital units.

References

1. Porter, M.E., Heppelmann, J.E.: How Smart, Connected Products Are Transforming Competition. Harvard Business Review 92, 64-88 (2014)
2. Yoo, Y., Boland, R.J., Lyytinen, K., Majchrzak, A.: Organizing for Innovation in the Digitized World. Organization Science 23, 1398-1408 (2012)
3. Downes, L., Nunes, P.F.: Big-Bang Disruption. Harvard Business Review 91, 44-56 (2013)
4. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.: Digital Business Strategy: Toward a Next Generation of Insights. MIS Quarterly 37, 471-482 (2013)
5. Hess, T., Barthel, P.: Wieviel digitale Transformation steckt im Informationsmanagement? Zum Zusammenspiel eines etablierten und eines neuen Managementkonzepts. HMD Praxis der Wirtschaftsinformatik 54, 313-323 (2017)
6. Porter, M.E., Heppelmann, J.E.: How Smart, Connected Products Are Transforming Companies. Harvard Business Review 93, 96-114 (2015)
7. Rigby, D.K., Sutherland, J., Takeuchi, H.: Embracing Agile - How to Master the Process that's Transforming Management. Harvard Business Review 94, 41-50 (2016)
8. Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M.: Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. MIS Quarterly 41, 223-238 (2017)
9. Sindemann, T., von Buttlar, H.: Konzerne auf den Spuren von Start-ups - Wie etablierte Unternehmen Innovation Labs, Acceleratoren und Inkubatoren als Instrument der digitalen Transformation nutzen. Infront Consulting & Management GmbH and Capital (2017)
10. Chanias, S., Hess, T.: Understanding Digital Transformation Strategy Formation: Insights from Europe's Automotive Industry. In: Proceedings of the 20th Pacific Asia Conference on Information Systems, pp. 1-15. (2016)
11. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for Formulating a Digital Transformation Strategy. MIS Quarterly Executive 15, 123-139 (2016)
12. Haffke, I., Kalgovas, B., Benlian, A.: The Role of the CIO and the CDO in an Organization's Digital Transformation. In: Proceedings of the 37th International Conference on Information Systems, pp. 1-20. (2016)
13. Singh, A., Hess, T.: How Chief Digital Officers Promote the Digital Transformation of their Companies. MIS Quarterly Executive 16, 1-17 (2017)
14. Horlach, B., Drews, P., Schirmer, I.: Bimodal IT: Business-IT Alignment in the Age of Digital Transformation. In: Proceedings of Mulitkonferenz Wirtschaftsinformatik (MKWI), pp. 1417–1428. (2016)
15. Jöhnk, J., Röglinger, M., Thimmel, M., Urbach, N.: How to Implement Agile IT Setups: A Taxonomy of Design Options. In: Proceedings of the 25th European Conference on Information Systems, pp. 1521-1535. (2017)
16. etventure, https://www.etventure.de/digital-lab/ (Accessed: 13.09.2018)
17. Siggelkow, N., Rivkin, J.W.: Speed and Search: Designing Organizations for Turbulence and Complexity. Organization Science 16, 101-122 (2005)
18. Besson, P., Rowe, F.: Strategizing Information Systems-Enabled Organizational Transformation: A Transdisciplinary Review and New Directions. The Journal of Strategic Information Systems 21, 103-124 (2012)
19. Venkatraman, N.: IT-Enabled Business Transformation: From Automation to Business Scope Redefinition. Sloan Management Review 35, 73-87 (1994)
20. Zimmermann, A., Raisch, S., Birkinshaw, J.: How Is Ambidexterity Initiated? The Emergent Charter Definition Process. Organization Science 26, 1119-1139 (2015)

21. March, J.G.: Exploration and Exploitation in Organizational Learning. Organization Science 2, 71-87 (1991)
22. O'Reilly III, C.A., Tushman, M.L.: Organizational Ambidexterity: Past, Present, and Future. Academy of Management Perspectives 27, 324-338 (2013)
23. Dixon, J.A., Brohman, K., Chan, Y.E.: Dynamic Ambidexterity: Exploiting Exploration for Business Success in the Digital Age. In: Proceedings of the 38th International Conference on Information Systems, pp. 1-17. (2017)
24. O'Reilly III, C.A., Tushman, M.L.: Organizational Ambidexterity in Action: How Managers Explore and Exploit. California Management Review 53, 5-21 (2011)
25. Hill, S.A., Birkinshaw, J.: Ambidexterity and Survival in Corporate Venture Units. Journal of Management 40, 1899-1931 (2014)
26. Randolph, W.A.: Matching Technology and the Design of Organization Units. California Management Review 23, 39-48 (1981)
27. Dougherty, D.: Organizing for Innovation in the 21st Century. In: Clegg, S.R., Hardy, C., Lawrence, T., Nord, W.R. (eds.) The SAGE Handbook of Organization Studies, pp. 598-617. SAGE Publications, Thousand Oaks, USA (2006)
28. Vaast, E., Levina, N.: Multiple Faces of Codification: Organizational Redesign in an IT Organization. Organization Science 17, 190-201 (2006)
29. Ramus, T., Velten, C.: Digital Labs – Ideenturbo der digitalen Transformation oder Statussymbol für Vorstände?. Managerkreis der Friedrich-Ebert-Stiftung (2016)
30. Nickerson, R.C., Varshney, U., Muntermann, J.: A Method for Taxonomy Development and its Application in Information Systems. European Journal of Information Systems 22, 336–359 (2013)
31. Yin, R.K.: Case Study Research - Design and Methods. SAGE Publications, Thousand Oaks, USA (2013)
32. Hanelt, A., Piccinini, E., Gregory, R.W., Hildebrandt, B., Kolbe, L.M.: Digital Transformation of Primarily Physical Industries – Exploring the Impact of Digital Trends on Business Models of Automobile Manufacturers. In: Proceedings of the 12th International Conference on Wirtschaftsinformatik, pp. 1313-1327. (2015)
33. Miles, M.B., Huberman, A.M., Saldana, J.: Qualitative Data Analysis - A Methods Sourcebook. SAGE Publications, Thousand Oaks, USA (2013)
34. Fichman, R.G., Dos Santos, B.L., Zheng, Z.E.: Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum. MIS Quarterly 38, 329-353 (2014)
35. Burgers, H., Jansen, J.J.P.: Organizational Ambidexterity and Corporate Entrepreneurship: The Differential Effects on Venturing, Innovation and Renewal Processes. Frontiers of Entrepreneurship Research 28, 1-16 (2008)
36. Jansen, J.J.P., Van Den Bosch, F.A.J., Volberda, H.W.: Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators. Management Science 52, 1661-1674 (2006)

Expectations vs. Reality – Benefits of Smart Services in the Field of Tension between Industry and Science

Simon Hagen[1] and Oliver Thomas[1]

[1] University of Osnabrueck, Chair of Information Management and Information Systems
Katharinenstr. 3, 49074 Osnabrueck, Germany
{simon.hagen,oliver.thomas}@uni-osnabrueck.de

Abstract. The term "Smart Service" gains increased interest in science and practice since it promises to significantly improve a company's value offering. However, its publicity might lead to overdrawn expectations, especially between practitioners and scientists working in this field. Therefore, we conduct a mixed-method study comparing the expected benefits of Smart Services in science and industry to help identify and close occurring gaps. The study consists of a literature review for the scientific point of view and a survey among practitioners to capture the benefits of Smart Services in both groups. The results predominantly reveal the same vision of both groups for Smart Services, but indicate slight but fundamental differences.

Keywords: Smart Service, Benefits, Literature Review, Survey, Industry

1 Introduction

Increasing digitalization of products and services resulted in "today's service world" [1] and the advent of Smart Services [2], which promise to significantly improve a company's service offerings and enable customized service features for customers through information and communication technology (ICT) [1]. Smart Services "raised high expectations on their potentials" [3] and are widely used to describe a company's innovativeness. In contrast, the research domain of Smart Services is still in its initial phase and after first strategies for its exploration were proposed [e.g. 1], it currently seeks for a foundational knowledge base [4]. In addition, a comprehensive insight of customer perceptions is required [1], but "only little is known about the level of user awareness and usage of smart services" [5].

Naturally, the lack of a theoretical foundation of Smart Services as well as missing methods and models for their implementation imposes challenges on both research and practice. On one hand, companies are aware of the topic, but on the other hand they just partly start implementing Smart Services and often fail or hesitate [6] due to their complexity. This results, amongst others, from the number of disciplines which have to be involved in their development and application [4] and leads to "limited empirical evidence […] about how the new markets for these novel services are created" [7]. In addition, different perceptions of Smart Services in science and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

practice in terms of their capabilities and benefits lead to gaps in expectations and thus weaken the joint work on the topic and an industry-wide introduction of Smart Services. Therefore, some scientific papers propose descriptions and structures for the topic [e.g. 8] or analyse barriers for an industry-wide adoption [9], but no common knowledge base in terms of definitions has been established so far. Thus, no structured development methods for such systems are present in practice and the systems cannot be classified accordingly. This can sometimes lead to the misuse of the term, e.g. for a marketing purpose [10].

With this publication, we address the gap between practitioner's expectations and scientific reality by conducting two separate studies, which compare the expectations of practitioners with benefits of Smart Services proposed in scientific literature. Thereby, our research fits into the research agendas of Wünderlich et al. [1] and Larson [4] and answers the following research question: *What are the main perceived benefits and expectations of Smart Services in science and practice and where do they resemble or differ?* The results help understand each other's perspective and thus bring both fields closer together, which is vital for a narrow exchange of knowledge and for an in-depth development of the topic. A common understanding enables an integrated view on the topic and consequently allow a higher value proposition in a shorter period of time. Especially the field of Information Systems (IS) can play a key role, since it aims at design-oriented and integrated development of systems at the interface between different disciplines [11, 12].

Therefore, after defining the term Smart Service (chapter 2), we conducted a structured literature review according to Webster & Watson [13] and identified 32 sources which mention benefits that result from implementing Smart Services in companies (chapter 4.1). In a second step, we conducted a web-based survey among 76 employees of different sized companies and asked them for their expectations by offering or using Smart Services (chapter 4.2). In addition, we had the participants validate the benefits identified in the literature as well (chapter 4.3). Finally, we compare the findings of the literature and the survey in chapter 5 and discuss the results in chapter 6.

2 Smart Services

The term "Smart Service" was primarily introduced to a scientific context by Allmendinger and Lombreglia in 2005 [2], but it has become popular in recent years. In the literature the term is explained with many different characteristics, e.g. awareness and connectivity [14] or its pre-emptive nature which is based on a "hard field intelligence" [2]. According to Wünderlich et al. [1] all characteristics fit into three groups: technological, customer oriented and context-specific perceptions. This highlights the nature of Smart Services, which relies on the use of IT to generate customer value according to current or upcoming situations.

Paluch [15] defines Smart Services therefore as "digital services, that are delivered through an intelligent and networked IT-infrastructure and generate value in combination with physical objects/products by continuous data collection and analysis". In a more generic definition Acatech [16] states that Smart Services are "individually configured bundles of (physical) products and services via the internet". This generic definition highlights the close relationship to Product-Service Systems (PSS), which also aim at fulfilling the customer needs in a holistic view, but do not focus on the technology [6, 17]. Some even state that PSS and Smart Services are the same thing [e.g. 18].

To achieve this holistic solution, Smart Services have to incorporate many different disciplines like engineering, computer science, economics or social sciences for development and application of the system [4]. They influence the business relationship of the provider and customer of a service and "are enabled and influenced by information that different industrial devices and processes generate, store and transmit to enable efficient operation, optimization, analysis and integration of business functions" [9]. Thus, Beverungen et al. [8] argue that Smart Services are enabled *by Smart Products, which* can be defined through core properties like unique identification, connectivity, sensors or actuators. The combination of different entities within a Smart Service is summarized by the term *Service System*, which refers to the same concepts according to Edvardson [19]. Spohrer et al. [20] define Service Systems as "socio-technical configurations of people, technologies, organizations, and information designed to create value by fulfilling the needs of those participating in the system" and Larsson [4] emphasizes the key role of the humans in the system. The National Science Foundation [21] adds in terms of the 'smart' component that a Smart Service System "amplifies or augments human capabilities to identify, to learn, to adapt, to monitor and to make decisions. The system does so through self-detection, self-diagnosing, self-correcting, self-monitoring, self-organizing, self-replicating, or self-controlled functions".

It becomes obvious that many terms are used in the same or likewise context and refer to similar concepts, which makes it difficult to distinct them correctly. Their concept has been applied in diverse industries like science and education, production industry, health sector or for individuals, e.g. in smart homes. This implies yet again the vast opportunities but also the complexity and the manifold obstacles in understanding, defining and developing them.

3 Method

To gain insights on both the scientific and practical perceived benefits of Smart Services we conducted a triangulation approach according to the mixed-method research concepts of Greene et al. [22]. Therefore, each method aims at identifying beneficial factors of Smart Services for the receiver or provider of the offering but from different perspectives. The findings of both studies, an extensive literature review and a web-based survey, are aggregated and discussed in section 5 and 6.

For the literature review we followed the approach of Webster and Watson [13] and used the databases *AISeL*, *EBSCOHost*, *ISI Web of Knowledge* and *SpringerLink* with the search term "smart service" in addition with the words advantage, benefit, value, result and their German counterparts. No further limitations regarding year of publication etc. were made, which led to 1.515 hits in total. After removing duplicates we filtered for title and abstract, which left us with 57 publications for detailed analysis. Interestingly, even though title and abstract sounded promising, less than half of these publications (25) contained useful insights on the beneficial factors of Smart Services. To obtain these, we applied two rules while analysing the publications: First, we only used publications that themselves rely on the term Smart Service. We did not evaluate and classify whether the descriptions in the paper fit to the definition of Smart Service, since there is no common understanding of the topic yet (c.f. Introduction) and therefore it is not clear which concept the authors were relying on. Second, we only considered benefits for companies using Smart Services which were stated as such, no deduction from an open statement (like a characteristic) into a benefit was made. This process was performed according to Mayring's [23] three-step process of content analysis (paraphrasing, generalizing, reducing) and led to the identification of 18 individual benefits, which could be grouped into five categories.

On the other hand, to analyse the perceived benefits of practitioners, we conducted a web-based survey among employees of German enterprises of various sizes. The survey was divided into three parts: (1) demographic data, (2) general concepts and (3) beneficial factors of Smart Services. To obtain unbiased insights on the opinions of the participants, we first asked them about their opinion of beneficial factors in a free text field (in part 3 of the survey). To get additional insights on their thoughts on the results from the scientific perspective, we secondly asked them about the factors identified in the literature review. We used two questions for each superordinate group (c.f. Section 4.1) to keep the survey in an adequate length. The questions had a 5-point Likert-Scale and were aligned to the guidelines of [24]. We chose this order to avoid distorted or biased results. In total 76 people participated, of which 40 fully completed the survey (52%).

4 Results of the Studies

4.1 Beneficial Factors of Smart Services in the Literature

The different understandings and shapes of Smart Services in the literature lead to various practical examples and therefore beneficial factors. A distinct trend in the topic can be identified when plotting the years in which the papers were published (c.f. Figure 1). By far most papers are from 2017 and this years (2018) number of publications is already (late summer) almost as high as 2005-2016 in total. In addition, we analysed the type of customer relationships (B2B or B2C) of Smart Services that were named in the publications. We did this by evaluating the given examples and for which customer group they are relevant. For each paper we did not count more than one occurrence of B2B or B2C group. As one can see in Figure 2, the term is applied in a B2B context nearly twice as much (19) compared to the B2C sector (11). Interestingly, some papers do not limit their scope to one sector and describe benefits for both.

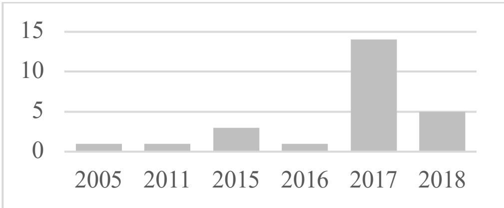


Figure 1. Publications per year in the literature review

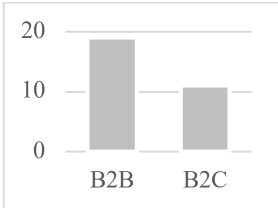


Figure 2. Mentioning's of customer segments

For the beneficial factors we identified 18 in total which we grouped into five segments (see Table 1 in the Appendix, chapter 8). In the following we will describe each group and the containing factors briefly. The whole matrix containing all publications and their concepts can be found in the appendix.

Group one includes benefits which in general describe effects of Smart Services in terms of **decreasing cost and time effort**. The costs are specified to 'traveling costs' (1 mention), since they are decreased due to opportunities of remote services, and 'search or transaction costs' (4 mentions), e.g. by using "fully automatic marketplaces" [16]. The majority (9 mentions) refers to 'reduced costs' in general, for instance through fewer downtime costs [25]. The 'reduction of time efforts' is usually explained as the receiver of a Smart Service has to spent less time in doing an action, since it is e.g. able to automate more tasks [26].

Group two summarizes benefits that in general **improve objects or conditions** through the application of Smart Services. In detail it specifies 'product' (5) and 'quality improvements' (6), 'increased performance' (12), 'higher flexibility' (7) and 'improved service provision' (2). While most of them are self-explanatory, the last one was described as enabling the customer of a Smart Service to benefit from a higher availability of the service providers offerings [1], which is in addition partly linked to flexibility, as the first enables the second.

Group three is named **customer related benefits**. Based on the definitions of Smart Services one could argue that they by nature add value to their customers, however some authors stated it explicitly. This group therefore focuses on the customer as a person or institution. It contains the factors 'relationship to the customer' (5), the 'customer experience' (5) which goes together with customer experience and a generic collection of 'values added for the customer' (12). Paluch [15] for example mentions the individual configurability, which implies an increased value for the customer. According to Allmendinger and Lombreglia [2] the relationship to a customer is "own[ed] […] as never before" by the offering company. The convenience that can accompany Smart Services can also lead to an increased customer experience [5]

Group four, **monitoring and maintenance**, is the second largest in terms of mentions (27 in sum), even though it contains only the two factors '(real time) monitoring' (17) and '(predictive) maintenance' (10). Monitoring is seen by some as a prerequisite [e.g. 27] and can be with regard to the machine itself, but also to observe the state of the object of interest, e.g. plants in agriculture [28]. In addition it not only enables maintenance, but for instance also adjusts the object (remotely) [8], which leads e.g. to increased performance (c.f. group two). Predictive maintenance enables providers e.g. to repair machines, which becomes especially valuable in production chains [6].

Table 1. Beneficial factors of Smart Services from the literature
(G=Group, M=Mentions, R=Relative Mentions, S=Sum of Mentions/Group)

G	*Benefit*	*M*	*R*	*S*
1) Cost/Time Reduction	Cost Reduction	9	7,63%	20
	Time Reduction	6	5,08%	
	Search/Transaction Cost Reduction	4	3,39%	
	Reduced Travel Cost	1	0,85%	
4) Monit. & maintenance	(Real time) Monitoring	17	14,41%	27
	(Predictive) Maintenance	10	8,47%	
3) Customer rel. benefits	Customer Value	12	10,17%	22
	Customer Relationship	5	4,24%	
	Consumer Experience	5	4,24%	

G	*Benefit*	*M*	*R*	*S*
2) Improvements	Increased Performance	12	10,17%	32
	Increased Flexibility	7	5,93%	
	Quality Improvement	6	5,08%	
	Product Improvement	5	4,24%	
	Increased Service Provision	2	1,69%	
5) Miscellaneous	Safety	5	4,24%	17
	Information Provision	5	4,24%	
	Environmental Benefits	5	4,24%	
	Partnerships	2	1,69%	

Group five, **miscellaneous**, holds beneficial factors that do not relate to any other group mentioned so far. It has the least mentions (17) of all groups and contains the factors ‘safety’ (5), ‘information provision’ (5), ‘environmental benefits’ (5) and ‘partnerships’ (2). Safety can be achieved in different contexts, e.g. in the industry [16], mobility [29] or in the health sector [30] by recognizing events and acting autonomously to prevent or decrease the harm. Through Smart Services information can e.g. be provided context-aware [31] or customized to the user [32]. Environmental benefits, e.g. fewer waste of scarce resources, can also be achieved through Smart Services [33]. Petrie et al. [34] argue that through more standardized services companies can partner in an easier and more efficient way.

4.2 Practitioners’ Expectations of Smart Services from a Web-Based Survey

In our web-based survey, we asked the participants in an open text field about their expectations and experiences with the benefits of Smart Services in their daily business. 76 practitioners participated of whom 40 finished the survey entirely. Most of the participants work at SMEs: 35% in companies with 101-200 employees, 17,5% have 11-100 colleagues and 10% work in very small firms (<10 employees). 37,5% are employed at big firms (<250 employees). They compete, amongst others, in industries like service (12,5%), mechanical engineering (10%) or forest and woodworking industry (17,5%). Of the 40 participants who completed the survey, 11 are female and 29 male. The majority (75%) is between 20 and 30 years old, the remaining 25% are older. 29 participants (72,5%) stated a job description that suggests intangible primary or supporting activities like clerk, marketing, sales or project manager. The remaining 20% work in jobs closely related to production or the products themselves, the remaining 7,5% cannot be classified (e.g. “student”).

The answers of the experts from the survey were clustered as well, which resulted in five groups: Alleviation of work, customer satisfaction, time/cost reduction, monitoring and maintenance and miscellaneous (see Table 2). As one can see, they in general go along with the clusters identified in the literature.

Table 2. Groups of beneficial factors of Smart Services from the web-based survey

#	*Group*	*No. of Coding’s (%)*	*Matching group from Literature*
1	Monitoring & Maintenance	31,0 %	4 - Predictive maintenance 23%
2	Assistance, Alleviation and Reduced Workload	21,4 %	2 - Improvements (partly) 27% 3 - Customer rel. benefits (partly)18
3	Customer Satisfaction	16,7 %	3 - Customer rel. benefits (partly)18
4	Reduction of Time and Cost	16,7 %	1 - Cost/Time Reduction 17%
5	Miscellaneous	16,7 %	None

The group with most mentions, **monitoring and maintenance**, compares to group four of the literature, which is the second most common there. In this context statements like “capture machine history”, “better/more precise monitoring” and even “predictive maintenance” were made, which are benefits for companies according to the interviewees. This is often linked to benefits resulting from them, e.g. “increased efficiency” of a machine or facility. In some cases also concrete data metrics were named which are valuable for the companies, for instance the feed speed of a machine, its ‘internal’ status or GPS coordinates. Interestingly only few remarks were made on the maintenance, the monitoring and capturing of data was mostly in the foreground.

Another major cluster deals with **assistance, alleviation and reduced workload** for the employees. Participants for example stated that they expect or experienced “simplified processes”, “improved information (flow)”, “less paperwork”, “decision support” and a “seamless production” by using Smart Services. This cluster closely relates to group three of the literature (customer related benefits) and shows that advantages do not only arise from an economic point of view.

Approximately the same amount of statements fit to the cluster **customer satisfaction**, which can be improved significantly by adapting Smart Services according to the participants. It partly also refers to group three of the literature (customer related benefits) and contains improvements for the customers like “better customer relationship” and “better customer satisfaction”, it enables them to provide “offers tailored more specific to customer needs” and to “optimize the services”.

The fourth group summarizing statements that indicate **reduction of time and cost** needed is also the fourth rarest mentioned one in the literature. The participants stated benefits such as “less downtime”, the “optimization of workflows” or an “increased efficiency”.

Lastly, remaining statements were grouped into **miscellaneous**. Despite sharing the same name with group 5 of the literature, in general no common statements were made. Benefits stated primarily refer to “fewer mistakes”, “more efficient machine utilization”, “improved customer acquisition” and “new business areas”.

4.3 Validation of Literature Findings from Practitioners in Web-Based Survey

In order to validate the findings of the literature we had the participants of the survey validate each group identified in the literature (c.f. Section 4.1). To prevent biasing their answers of their own expectations the validation took place at the end of the survey. In addition, to keep the survey short and prevent the participants from interrupting, we had two questions per group summarizing the findings of the literature. Each questions was answered on a 5-point Likert-Scale, asking for the consent (1:“Strongly Agree”) or refusal (5: “Strongly disagree”) of the statement. We left out group five (miscellaneous), since it cannot be summarized accordingly.

The following table shows the average of all given answers and the standard deviation for both questions.

Table 3. Results of the evaluation of benefits of Smart Services from the literature review (c.f. Section 4.1)

		Group 1	Group 2	Group 3	Group 4
Question 1	**Avg.**	2,25	2,13	2,18	2,15
	Std. Dev.	0,86	0,98	0,63	0,88
Question 2	**Avg.**	2,73	2,23	1,98	1,93
	Std. Dev.	1,30	0,88	0,72	0,75

5 Intersections of Expectations and Benefits

In general, the clusters from the literature review and the survey are quite similar in terms of the underlying concepts. However, the answers of practitioners are more specific or concrete and less abstract, probably since they have own experiences or ideas how Smart Services can benefit their company. This infers that both groups are on a similar track regarding their ideas of Smart Services, but on different levels of abstraction.

One topic that stands out is “predictive maintenance”. It ranks second in terms of mentioning’s in the literature and in the survey the group “monitoring & maintenance” is the most mentioned one. This implies that the connection to a physical product seems to be a core element of Smart Services, since it allows an in-depth observation of the object itself or other elements of interest in their surroundings. Besides real-time information this enables forecasting’s or improvements based on statistics and thus becomes essential for the whole system.

The cluster “assistance, alleviation and reduced work” from the survey (second most mentions) refers to the topics mentioned in groups two (improvements) and three (customer related benefits) from the literature. This is probably due to the different points of view, since the practitioners experience the effects of Smart Services as alleviations and in the literature they are described objectively. However, the "customer satisfaction" was mentioned as a single group in the survey as well, but only matches some of the "customer related benefits" from the literature.

The "reduction of time and costs" was mentioned similarly in the literature and from the practitioners, but only ranks in the lower third in terms of relative number of mentioning's. This group is the only exception to the above mentioned differentiating levels of abstractions between literature and practice. The statements by the practitioners, which otherwise are concrete and contain specific expectations, are more generic in this group. This infers, that the factor is assumed to be prevalent and not a core benefit to practitioners, even though it is a crucial aspect for a widespread dissemination of Smart Services.

The last group, miscellaneous, cannot be compared in a reasonable way, since the containing topics are too different. Also, each domain stated benefits that were not mentioned by the other at all, e.g. 'safety' or 'environmental benefits'. Those were listed in the literature, but not mentioned by the practitioners. Probably due to their increased relevance in the current phase of Smart Service dissemination and adoption.

The validation of the benefits from the literature in the survey (c.f. Section 4.3) mostly confirms these findings and the similarities, since each question (except Group 1, Question 2) scores a value about 2, which refers to "agree" on a 5-point Likert-Scale. The groups with the best scores in average are "monitoring & maintenance" (4) and "time and cost reduction" (3), which exactly matches the results described above. The evaluations of "alleviation of work" (1) and "customer satisfaction" (2) also point out their relevance, but are not that important.

6 Discussion, Limitations and Outlook

The term "Smart Service" gains increased interest in science (c.f. Figure 1) and in practice many companies acknowledge the benefits that arise from implementing them. However, the expectations and possibilities from both domains seem to differ and no methods for concrete adoption are present. That is why companies hesitate to implement them in full regard. Thus, we compared the benefits of Smart Services stated in science and practice, which originate from an extensive literature review and a web-based survey with experts from companies. This enabled us to answer our research question in a two step process: The first part of the question ("What are the main perceived benefits and expectations of Smart Services in science and practice […]") is answered in sections 4.1 and 4.2 and allows the comparisons of the findings ("[…] and where do they resemble or differ?") in section 5. In this section we discuss the results in the context of IS-research and derive steps for future research to overcome the identified gaps and to build upon common ideas and experiences.

The findings described in the previous sections, the consensus for benefits of Smart Services between science and practice, appear to be obvious in a first instance. This does not seem to be unexpected, since the term "Smart Service" is attracting increased interest and the ideas are exchanged between the domains. The differences are predominantly on the different levels of abstraction. For companies the concrete implementation of a Smart Service is most important, thus they derive benefits from the offerings they can supply so far or know about. In a scientific context however the scope is, due to the more generic approach, broader and thus more widespread ideas arise, e.g. the opportunities of a positive impact on ecologic sustainability. However, the practitioners also mention several ideas that have not yet been mentioned in a scientific context. Those ideas are assorted in group 5 five (miscellaneous) and do not match any group or even concrete statement from the literature. This implies that there are still topics that remain uncovered by one of the groups evaluated (science and practice) and leaves room to exchange experiences and ideas in both directions.

Both domains agree that Smart Services have the potential to fundamentally improve or change the way companies generate value for their customers, that's probably the reason why this is the largest group in terms of mentioning's in the literature (32 in total, group two). It is interesting that one of the core elements in both literature and practice is the strong connection to physical products, expressed e.g. by the expected functionalities and benefits of predictive maintenance. Even though only a few participants of the survey are from the machinery industry and no focus or hint on this topic was given in the survey, it is the most mentioned aspect. This supports one of the current streams in IS research, which proposes such strong connection between the physical and intangible components [e.g. 8]. However, substantial adoptions are scarce, which seems to be similar to the adoption of PSS after their introduction in a scientific context. The concepts and benefits of PSS are widely agreed on in science and practice, but practical methods for their development are still not prevalent. In addition, many benefits named by the practitioners, except predictive maintenance, could also be realized by applying "regular PSS" without a "smart component". This verifies the unclear separation of PSS and Smart Services in both domains but at the same time confirms their strong interconnectedness (c.f. Section 2). To simplify adoption in practice it seems to be helpful to highlight the relations and point out the underlying concepts in order to build upon existing knowledge and implementation initiatives. The analysis of the progression from PSS to Smart Services and the development of Smart Services themselves through various scientific iterations can support the adoption in practice as well, by enabling companies to adept the steps of development from science.

Another interesting result is that there is no distinct separation of benefits from participants of the survey working in "intangible" or "tangible" jobs (c.f. Section 4.2). Both groups share the same ideas in terms of product, service or business related features. The same applies for examining the answers across different industries. The participants seem to have rather comprehensive ideas of Smart Services and their concepts and benefits and are not limited to their own perspective. However, we do not find a correlation between the groups. Scientists can build upon this broad view to analyse the often complex structures of Smart Service (Systems) [20]. This goes along with the finding that practitioners from many different disciplines propose beneficial factors of Smart Services despite their relation to information or engineering research. Therefore, from an IS-research perspective, it seems to be valuable to cooperate with other research disciplines in this field, e.g. product or production engineering [4]. This would also accommodate to the strong relationship of the physical and intangible parts. In addition, a common concept or technology for an in-depth integration of the various disciplines seems to be valuable.

To foster the cooperation of science and practice and to allow sharing of knowledge between the disciplines, we can derive the following future research approaches. Practitioners gave many examples of Smart Service applications, but only few mentioned their importance on generating valuable business model. Scientists can support this change in mind-set by developing frameworks to structure and implement sustainable value offerings. Also, our results reveal benefits which are not common yet, e.g. a building relationships between companies through Smart Service or an improved environmental impact. Research can focus on these benefits to extend the scope of Smart Services beyond the smart product and service perspective solely. The investigation of Smart Service in the pure service sector, e.g. banking, is another topic that needs further consideration in order to prove if the value of Smart Services only realizable in combination with physical products or how the concepts can be transformed and adopted. This matches with the possibilities of substituting the physical and service components in the system and the question, how this affects the conceptualization of Smart Services from a research perspective and the value offering for companies.

This contribution serves as a starting point for future studies of Smart Services and therefore has some minor limitations. The survey does not allow empirical evaluations, since it has not enough participants to guarantee for statistical evidence and practitioners from more industry sectors would improve the results. Participants from “intangible” jobs, which might bias the findings, also dominate the present group. In addition, the validation of the clusters found in the literature with only two questions each is just small indication for its validity. Like every other content analysis (literature and survey), the coding of the benefits and grouping them is a partly subjective task, particularly when the benefits often determine and result in each other. Finally, another data source for insights on practice, like white papers, could have been used.

In conclusion, many streams of Smart Services are similar in science and practice and, even though they are often on different levels of abstraction, they proceed in similar directions. For many research topics stated in prior scientific publications proposals for further investigation were already made, however they need to incorporate the concrete requirements and ideas from practice, to prevent divergence of the perspectives.

7 Acknowledgements

This research is partly funded by the European Regional Development Fund and the State of Lower Saxony (NBank) in the scope of the research project SmartHybrid – Service Engineering (ID: 6-85003236). We would like to thank them for their support.

References

1. Wünderlich, N. V., Heinonen, K., Ostrom, A.L., Patricio, L., Sousa, R., Voss, C., Lemmink, J.G.A.M.: "Futurizing" smart service: implications for service researchers and managers. J. Serv. Mark. 29, 442–447 (2015).
2. Allmendinger, G., Lombreglia, R.: Four Strategies for the Age of Smart Services. Harv. Bus. Rev. 10, 131–145 (2005).
3. Biehl, S.: Design guidelines for smart services: A strategic-logic perspective on seeking competitive advantage with digitized servitization strategies, (2017).
4. Larson, R.C.: Smart Service Systems: Bridging the Silos. Serv. Sci. 8, 359–367 (2016).
5. Peng, G.C.A., Nunes, M.B., Zheng, L.: Impacts of low citizen awareness and usage in smart city services: the case of London's smart parking system. Inf. Syst. E-bus. Manag. 15, 845–876 (2017).
6. Kampker, A., Frank, J., Schwartz, M., Jussen, P.: Lernen von den Besten: Fünf Erfolgsfaktoren bei der Entwicklung von Smart Services. In: Service Engineering. pp. 151–165. Springer Fachmedien Wiesbaden, Wiesbaden (2018).
7. Hu, Y., McLoughlin, D.: Creating new market for industrial services in nascent fields. J. Serv. Mark. 26, 322–331 (2012).
8. Beverungen, D., Müller, O., Matzner, M., Mendling, J., vom Brocke, J.: Conceptualizing smart service systems. Electron. Mark. 1–12 (2017).
9. Töytäri, P., Turunen, T., Klein, M., Eloranta, V., Biehl, S., Rajala, R., Hakanen, E.: Overcoming Institutional and Capability Barriers to Smart Services Methodology. Hawaii Int. Conf. Syst. Sci. 2017. 1642–1651 (2017).
10. Duddy, K.: What would smart services look like: And how can we build them on dumb infrastructure? Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 5472 LNCS, 5–14 (2009).
11. Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J.: Memorandum on design-oriented information systems research. Eur. J. Inf. Syst. 20, 7–10 (2011).
12. Mertens, P., Barbian, D.: Researching "Grand Challenges." Bus. Inf. Syst. Eng. 57, 391–403 (2015).
13. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, 13–23 (2002).
14. Maglio, P.P.: Editorial Column—Smart Service Systems. Serv. Sci. 6, i–ii (2014).
15. Paluch, S.: Smart Services – Analyse von strategischen und operativen Auswirkungen. In: Dienstleistungen 4.0. pp. 161–182. Springer Fachmedien Wiesbaden, Wiesbaden (2017).
16. Acatech: Smart Service Welt - Umsetzungsempfehlung für das Zukunftsprojekt Internetbasierte Dienste für die Wirtschaft. , Berlin (2015).
17. Beverungen, D., Matzner, M., Janiesch, C.: Information systems for smart services. Inf. Syst. E-bus. Manag. 15, 781–787 (2017).
18. Anke, J.: Design-integrated financial assessment of smart services. Electron. Mark. (2018).
19. Edvardsson, B., Enquist, B.: The service excellence and innovation model: Lessons from IKEA and other service frontiers. Total Qual. Manag. Bus. Excell. 22, 535–551 (2011).
20. Spohrer, J., Maglio, P.P., Bailey, J., Gruhl, D.: Steps toward a science of service systems. Computer (Long. Beach. Calif). 40, 71–77 (2007).
21. National Science Foundation: Partnerships for Innovation: Building Innovation Capacity, https://www.nsf.gov/pubs/2016/nsf16591/nsf16591.pdf.
22. 22. Greene, J.C., Caracelli, V.F., Wendy, F.: Toward a Conceptual Framework for Mixed-Method Evaluation Designs. Educ. Eval. Policy Anal. 11, 255–274 (1989).

23. Mayring, P., Brunner, E.: Qualitative Inhaltsanalyse Der Stellenwert der Inhaltsanalyse im Methodenspektrum. Qual. Marktforsch. Konzepte - Methoden - Anal. 669–680 (2007).
24. Porst, R.: Frageformulierung. In: Baur, N. and Blasius, J. (eds.) Handbuch Methoden der empirischen Sozialforschung. pp. 687–699. Springer Fachmedien Wiesbaden, Wiesbaden (2014).
25. Pöppelbuß, J., Durst, C.: Smart Service Canvas - Ein Werkzeug zur strukturierten Beschreibung und Entwicklung von Smart-Service-Geschäftsmodellen. In: Bruhn, M. and Hadwich, K. (eds.) Dienstleistungen 4.0. pp. 92–111. Springer Gabler, Wiesbaden (2017).
26. Bullinger, H.-J., Ganz, W., Neuhüttler, J.: Smart Services – Chancen und Herausforderungen digitalisierter Dienstleistungssysteme für Unternehmen. In: Dienstleistungen 4.0. pp. 97–120. Springer Fachmedien Wiesbaden, Wiesbaden (2017).
27. Dreyer, S., Olivotti, D., Lebek, B., Breitner, M.H.: Towards a Smart Services Enabling Information Architecture for Installed Base Management in Manufacturing. 13th Int. Conf. Wirtschaftsinformatik. 31–45 (2017).
28. Maaß, W., Pier, M., Moser, B.: Smart Services in der Landwirtschaft. In: Service Engineering. pp. 167–181. Springer Fachmedien Wiesbaden, Wiesbaden (2018).
29. Peters, C., Maglio, P., Badinelli, R., Harmon, R.R., Maull, R., Spohrer, J.C., Tuunanen, T., Vargo, S.L., Welser, J.J., Demirkan, H., Griffith, T.L., Moghaddam, Y.: Emerging Digital Frontiers for Service Innovation. Commun. Assoc. Inf. Syst. 39, 136–149 (2016).
30. Gimpel, H., Huber, R.X.R., Röglinger, M., Hosseini, S., Probst, L., Faisst, U., Huber, R.X.R., Probst, L., Röglinger, M., Faisst, U.: Structuring Digital Transformation: A Framework of Action Fields and its Application at ZEISS. J. Inf. Technol. Theory Appl. 19, 31–54 (2018).
31. Casillo, M., Colace, F., Santo, M. De, Lemma, S., Lombardi, M.: A Context-Aware Mobile Solution for Assisting Tourists in a Smart Environment. Hawaii Int. Conf. Syst. Sci. 2017. (2017).
32. Watanabe, K., Mochimaru, M.: Expanding Impacts of Technology-Assisted Service Systems Through Generalization: Case Study of the Japanese Service Engineering Research Project. Serv. Sci. 9, 250–262 (2017).
33. Demirkan, H., Bess, C., Spohrer, J., Rayes, A., Allen, D., Moghaddam, Y.: Innovations with smart service systems: Analytics, big data, cognitive assistance, and the internet of everything. Commun. Assoc. Inf. Syst. 37, 733–752 (2015).
34. Petrie, C., Hochstein, A., Genesereth, M.: Semantics for Smart Services. In: Demirkan, H., Spohrer, J.C., and Krishna, V. (eds.) The Science of Service Systems. pp. 91–105. Springer, Boston, MA (2011).
35. 35. Huang, T.-L.: Creating a commercially compelling smart service encounter. Serv. Bus. 12, 357–377 (2018).
36. Cedeño, J.M. V, Papinniemi, J., Hannola, L., Donoghue, I.: Developing smart services by internet of things in manufacturing business. LogForum. 14, 59–71 (2018).
37. Nezhad, H.R.M., Schwartz, L.: Towards Open Smart Services Platform 2 . Service Consumption : A Changing. Hawaii Int. Conf. Syst. Sci. 2017. 1103–1109 (2017).
38. Pourzolfaghar, Z., Helfert, M.: Taxonomy of Smart Elements for Designing Effective Services. AMCIS 2017 Proc. 1–10 (2017).
39. Wiegard, R.B., Breitner, M.H.: Smart services in healthcare: A risk-benefit-analysis of pay-as-you-live services from customer perspective in Germany. Electron. Mark. 1–17 (2017).
40. Jüttner, U., Windler, K., Schäfer, A., Zimmermann, A.: Design von Smart Services – Eine explorative Studie im Business-to-Business-Sektor. In: Dienstleistungen 4.0. pp. 335–361. Springer Fachmedien Wiesbaden, Wiesbaden (2017).

8 Appendix

Table 4. Concept matrix of beneficial factors of Smart Services in the literature (mapping to benefit-names in table beneath)

Source	Sum	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
(Acatech 2015)	9	x	x	x	x					x	x	x	x	x					
(Kampker et al. 2018)	4	x	x							x		x							
(Allmendinger and Lombreglia 2005)	7	x	x		x	x				x	x	x							
(Biehl 2017)	5	x					x				x						x	x	
(Casillo et al. 2017)	2												x			x			
(Petrie et al. 2011)	4						x				x	x							x
(Beverungen et al. 2017)	8	x	x				x				x	x			x		x	x	
(Demirkan et al. 2015)	8		x		x		x		x		x				x	x	x		
(Dreyer et al. 2017)	1		x																
(Gimpel et al. 2018)	6		x	x		x	x					x							x
(Bullinger et al. 2017)	7	x	x			x	x				x	x					x		
(Huang 2018)	3											x	x	x					
(Cedeño et al. 2018)	4		x							x	x				x				
(Nezhad and Schwartz 2017)	1													x					
(Peng et al. 2017)	5		x							x			x	x	x				
(Peters et al. 2016)	2			x							x								
(Pourzolfaghar and Helfert 2017)	2		x												x				
(Pöppelbuß and Durst 2017)	5	x	x		x					x								x	
(Paluch 2017)	10	x	x	x		x				x	x	x	x			x	x		
(Wiegard and Breitner 2017)	3									x		x						x	
(Töytäri et al. 2017)	3	x	x									x							
(Jüttner et al. 2017)	4		x							x	x					x			
(Watanabe and Mochimaru 2017)	5										x	x				x	x	x	
(Maaß et al. 2018)	2		x															x	
(Wünderlich et al. 2015)	8	x	x	x	x	x	x	x	x										
Sum	**118**	**10**	**17**	**5**	**5**	**5**	**7**	**1**	**2**	**9**	**12**	**12**	**5**	**4**	**5**	**5**	**6**	**6**	**2**

1	(Predictive) Maintenance	10	Increased Performance
2	(Real time) Monitoring	11	Customer Value
3	Safety	12	Consumer Experience
4	Product Improvement	13	Search/Transaction Cost Reduction
5	Customer Relationship	14	Environmental Benefits
6	Increased Flexibility	15	Information Provision
7	Reduced Travel Cost	16	Time Reduction
8	Increased Service Provision	17	Quality Improvement
9	Cost Reduction	18	Partnerships

Innovation Networks and Digital Innovation: How Organizations Use Innovation Networks in a Digitized Environment

Axel Hund[1,2], Heinz-Theo Wagner[2]

[1] University of Bamberg, Chair of Information Systems and Services, Bamberg, Germany
{axel.hund}@uni-bamberg.de
[2] German Graduate School of Management & Law, Management and Innovation, Heilbronn, Germany
{heinz-theo.wagner}@ggs.de

Abstract. The digital transformation dramatically lowered the costs for communication and coordination, thus, enabling new forms of cooperation. Companies seize this opportunity by creating new types of innovation networks. Until now, we know little about which types of innovation networks are currently prevalent and why organizations use them. In this paper, we build upon a recent study dealing with categorization of innovation networks and present the results of an exploratory series of case studies conducted with 27 high-level executives from 11 organizations in various industries. Our results indicate that companies are maintaining high-levels of centralized control over the innovation network, which is contrary to what the literature suggests. Furthermore, there is a strong trend towards more heterogeneous knowledge within a network. Additionally, we identify mechanisms that help companies to transition from one type of innovation network to another one and investigate why organizations use certain innovation networks.

Keywords: innovation network, digital innovation, digital transformation, case study research

1 Introduction

'Two heads are better than one' is a common proverb and the increasing reliance on innovation networks - instead of individual innovators - appears to confirm this assumption (e.g., [1],[2]). Innovation networks can be understood and defined as "a set of actors connected by a set of ties. The actors […] can be persons, teams, organizations, concepts, etc." [3]. For this paper, we have a focus on socio-technical networks formed between various actors and their respective tools in an increasingly digitized environment as described by Lyytinen et al. [1].

The increasing reliance on different forms of digitally enabled innovation networks is reflected in widespread practices such as open innovation [4], user innovation [5] and crowdsourcing [6]. These practices are enabled by the progressive digitization of our environment (e.g., [7],[8]) with costs for communication and computation at an unprecedented low level [9], which, enables companies to easily access knowledge from

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

beyond their organizational boundaries [10], [11]. Hence, innovation networks are poised to become the backbone of successful innovation efforts in a digitized environment [1], [2], [12]. Accordingly, recent research addresses the topic of innovation networks. For example, striving for conceptual clarity Lyytinen et al. [1] built upon extant literature on innovation networks and put forward a framework to distinguish between different forms of innovation networks by categorizing them along the dimensions of (1) control over resources and (2) degree of heterogeneity.

However, even though we can observe the rise of innovation networks over single innovators (e.g., [4], [5], [10]), and know that different types of innovation networks lead to either incremental or radical innovation outcomes [1], we do not know which types of innovation networks are actually prevalent among incumbent companies since digitization became nearly ubiquitous. Furthermore, we do not know how companies can fluidly transition between different types of innovation networks. Moreover, we do not know why organizations decide to use a certain innovation network. Thus, as pointed out by Nambisan et al. [12] we cannot comprehend "how firms are able to successfully and fluidly mix, match, and integrate internal and external parties and various diverse communities in digital innovation". As innovation is the ultimate raison d'être for companies [13] it is paramount to further examine how the pervasive digitization of entire industries influences innovation networks and why organizations use specific network types. Thus, this paper investigates two research questions:

RQ1: Which types of innovation networks are currently prevalent and why do organizations use them?

RQ2: How can companies fluidly transition between different types of networks?

Striving to answer these questions, we conducted 27 interviews with high-level managers from 11 different companies in different industries and inquired about their approaches to organizing and innovating in a digitized environment. Theoretically, we rely on the framework by Lyytinen et al. [1] to clearly delineate and categorize different forms of innovation networks. This allows us to distinguish between common forms of innovation networks, categorize them within the framework and examine which types of innovation networks are currently prevalent and why companies use them. Additionally, it allows us to identify mechanisms that companies use to transition from one type of innovation network to another. In the following sections, we will review the literature on the influence of digitization on innovation networks and describe how we analyzed the case studies. Furthermore, we present our findings, discuss the implications of our findings and highlight potential limitations and promising avenues for future research.

2 Related Literature

New types of innovation networks are emerging as traditional approaches struggle to efficiently coordinate the increasing amount of connected actors and integrate the vastly heterogeneous knowledge available [1]. There are entire literature strands that take a more granular look on different types of innovation networks such as network-centric innovation [2], open innovation [4], [14], distributed innovation [15], user innovation [5], [16] and crowdsourcing [6], [17]. All of those literature strands have in common that they accentuate an external focus for innovation. This makes perfect sense when considering that "[…] in any given sphere of activity most of the pertinent knowledge will reside outside the boundaries of any one organization, and the central challenge for those charged with the innovation mission is to find ways to access that knowledge" [15]. Thus, organizations experience an increasing blurring of their external boundaries and leverage digital technology to enable new cooperation's [12]. All of the beforehand mentioned articles focus on different aspects of external cooperation. For example, user innovation focuses on the involvement of the user in the development process to better understand user needs [5]. Crowdsourcing on the other hand is about outsourcing tasks to an undefined crowd through tournament-style or collaboration-style open calls [6]. In addition, digital technology also plays a vital role for more traditional approaches such as merger and acquisitions (M&A), as digital technology-related M&As can make or break the efforts of a company to master the digital transformation [18], [19]. In general, research on networks has grown exponentially and can be categorized on basis of the respective focal point such as for example social capital, embeddedness or organizational networks [3]. Another typology put forward by Mentzas et al. [20] delineates inter-organizational knowledge networks along the nature of exchange (sharing / trading) and the nature of community (closed / open). However, since our research focuses on *innovation* networks in a digitized environment, we chose to build upon the theoretical framework put forward by Lyytinen et al. [1] as it was designed to do exactly that. The framework helps delineate between different types of innovation networks along two dimensions. These two dimensions embody the effect of ubiquitous digital technology, which leads to:

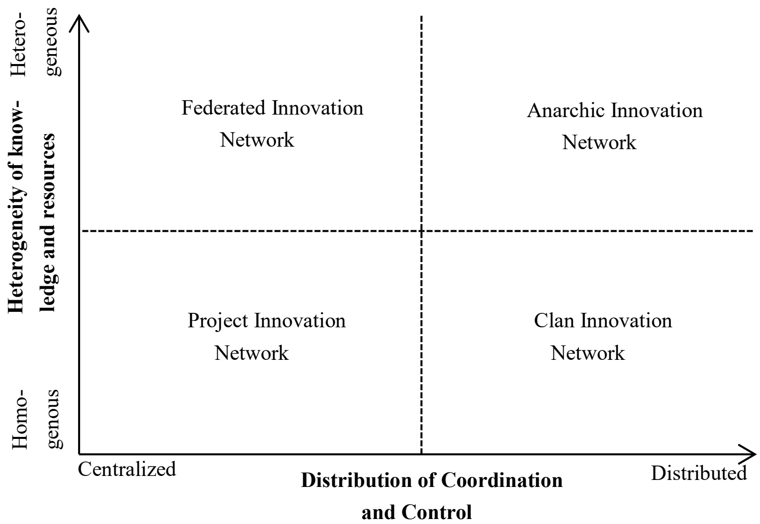
(1) **Distribution of coordination and control:** Vanishing costs for communication allow connecting, coordinating and controlling innovation network contributors and resources independently of time and location. In the past, business models and innovation networks were limited by the high cost of communicating and processing information [21]. However, with the rise of ubiquitous digitization these costs have diminished to the brink of almost vanishing [9]. Plummeting costs for communication make it possible to connect and coordinate previously unconnected actors and, thus, enabled the creation of new types of innovation networks (e.g., [4], [6]).

(2) **Heterogeneity of knowledge and resources:** Combining the expertise of previously separated knowledge communities increases the heterogeneity of resources and knowledge within the innovation network. When a formerly static product or process is digitized it is endowed with an unprecedented level of flexibility and openness [8], [22]. Increased openness and flexibility enables previously separated knowledge communities to combine their distinct areas of expertise as demonstrated in the "quadruple-play" – the combination of phone functionalities with TV-services, broadband internet and mobile internet apps [1], [8]. The convergence of heterogeneous knowledge enabled the creation of a groundbreaking digital innovation - the smartphone [1], [8]. As different knowledge communities are more interconnected and form innovation networks, the heterogeneity of knowledge available within these innovation network increases [1]. Moreover, due to the higher flexibility, contributors can frequently enter and exit the innovation network when the focus of an innovation network changes or

requires different forms of knowledge [23]. This dynamism increases the heterogeneity even further as the contributors in innovation network constantly change [12].

Lyytinen et al. [1] conceptualize along those two dimensions four canonical archetypes of innovation networks which can be intra-organizational or inter-organizational. The first archetype is *project innovation networks* which is characterized by homogenous contributors coming from the same discipline and, therefore, have the same knowledge background. Project innovation networks rely (mostly) on hierarchically integrated structures as given within an organization, thus, allowing for tight control of resources and the goal of the innovation efforts [1]. The second archetype is *clan innovation networks,* which also comprises a homogenous group of contributors that have the same – or at least very similar – knowledge background. However, in a clan innovation network the contributors are not controlled by a hierarchical structure and can freely determine which innovation outcome they pursue. A typical example of a clan innovation networks are open-source communities [1]. The third archetype is *federated innovation networks,* which are characterized by contributors from different disciplines and heterogeneous knowledge. Even though they come from different backgrounds, resources and the eventual innovation outcome are tightly controlled by hierarchically integrating the contributors. Classic examples of a federated innovation network are automotive manufacturers, which rely on numerous knowledge communities ranging from logistics over engineering to design. Contributors from different departments or even different companies work together beyond the classic company or discipline-boundaries. Nonetheless, the final outcome of innovation efforts is tightly controlled by the manufacturer [1], [24]. Lastly, there are *anarchic innovation networks,* which are characterized by contributors from different disciplines with heterogeneous knowledge backgrounds and no hierarchical control. Hence, anarchic innovation networks introduce an unprecedented level of complexity and dynamism as vastly heterogeneous knowledge communities' work together in absence of any formalized form of control over the innovation outcome, the innovation process or even the structure of the network. An example for an anarchic innovation network was the Gehry project in Bilbao in which numerous independent companies from diverse backgrounds took part [25]. Figure 1. depicts the four types of innovation networks along two dimensions:

Figure 1. Four Types of Innovation Networks



3 Method

“Case studies are the preferred strategy when "how" or "why" questions are being posed, when the investigator has little control over events, and when the focus is on a contemporary phenomenon within some real-life context” [26]. Due to the nature of our research, we neither have control over the events nor the context, and are fundamentally guided by “how” and “why” research questions. Thus, case study research is a perfect match [26]. For the sampling procedure, we focused on companies that currently conduct digitalization projects and aimed a minimum of three senior managers per company to ensure multi-facetted insights into the project. However, depending on company size, sometimes one person was in charge of every critical area of interest. In total, we interviewed 27 senior managers from 11 different companies in various industries as listed in table 1:

Table 1. Case Study Overview: Industry, ID, Interviewee Position and Length

Case	Industry	ID	Interviewee Position	Length
A	Mechanical Engineering	IP 01	Innovation Manager	59 min
B	Banking	IP 02	Innovation Manager	68 min
		IP 03	Head of Product Management	66 min
		IP 04	Chief Technical Officer	57 min
C	Online Publishing	IP 05	Head of Business Development	72 min
		IP 06	Deputy General Manager	70 min
		IP 07	Chief Technical Officer	66 min
D	Private Bank	IP 08	Marketing Manager	72 min
		IP 09	Head of IT & Organization	83 min
		IP 10	Chief Executive Officer	45 min
E	Plant Construction	IP 11	Product Group Manager	91 min
		IP 12	Head of Automation and Controls	121 min
		IP 13	Director Technology Management	54 min
F	Banking	IP 14	Chief Digital Officer	57 min
G	Banking	IP 15	Chief Executive Officer	65 min
H	Building Industry Supplier	IP 16	Chief Digital Marketing Manager	53 min
		IP 17	Chief Financial Officer	46 min
		IP 18	Chief Marketing Manager	53 min
		IP 19	Chief Executive Officer	45 min
I	Banking	IP 20	Chief Digital Officer	60 min
J	Automotive and Aircraft Supplier	IP 21	Vice President (VP) Sales	63 min
		IP 22	Product Manager	33 min
		IP 23	Team Lead Electronics	109 min
K	Automotive Manufacturer	IP 24	VP Systems Development	70 min
		IP 25	VP IT Solutions	60 min
		IP 26	Director Corporate Strategy	35 min
		IP 27	Director Product Management	45 min

The interviewed senior managers held positions in various departments such as IT, innovation, and R&D. Most interviews were conducted during 2016 and onsite and were guided by the following guidelines: First, we asked the interviewees to briefly summarize their background and to describe their position in the company. Subsequently, we asked how they personally define digitization and whether they could talk about a recent digital initiative. Following up, we asked in detail about the company's internal structures and decision hierarchies. Furthermore, we asked about innovation processes, innovation co-operations and resource allocations. Additionally, we included questions concerning the personal opinions of the interviewed senior managers to gain a better understanding of critical success factors. The interviews were recorded and subsequently transcribed. The qualitative data analysis was conducted deductively following the guidelines of Mayring and Fenzl [27]. We started by defining clear research questions and, henceforth, selected and prepared the appropriate material as described above. We then searched the literature for frameworks that help determine categories which are firmly grounded in theory – in our case Lyytinen et al. [1] provided an ideal place to start as the framework specifically focuses on innovation networks in a digitized environment. Subsequently, guided by existing best practices in the literature (e.g., [27], [28]) we deductively coded the interviews in accordance with our established coding guideline as depicted in table 2:

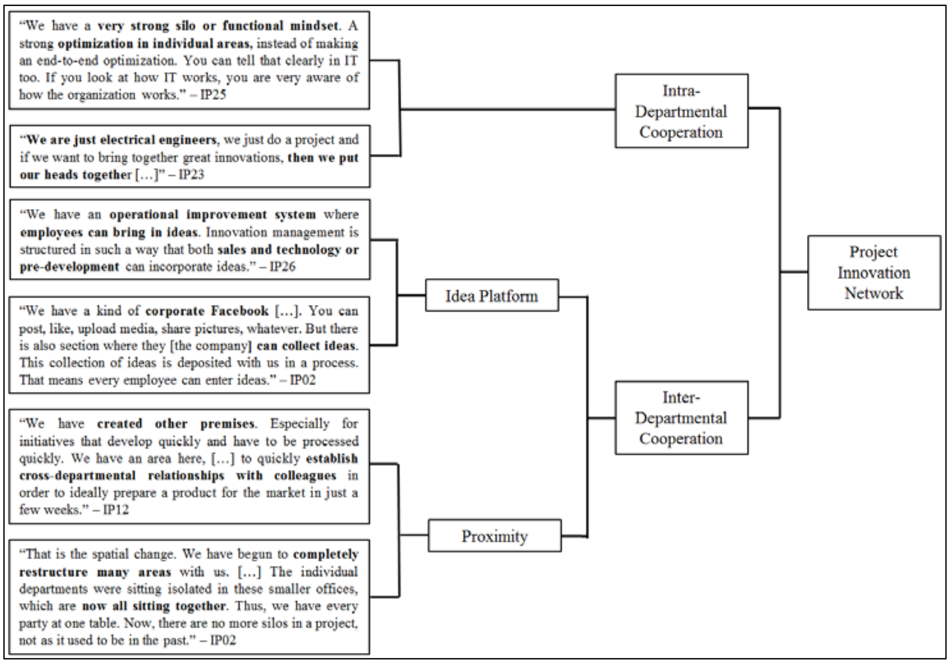
Table 2. Coding Guidelines based on Lyytinen et al. [1]

Type	Network Characteristics	Coding Rules
Project	**I.** Homogeneous actors / tools, readily identified and mobilized **II.** Hierarchically integrated control, mostly within a single company	There are clearly identifiable control structures for resources and outcomes. Actors within the network must have homogeneous knowledge and are easily identifiable.
Clan	**I.** Homogeneous pool of dynamic actors driven by common interest and well-defined set of tools; readily identified & mobilized **II.** No centralized hierarchical control	No clearly identifiable control structure. Outcome control is not existent or distributed. Actors within the network must be easily identifiable and have homogeneous knowledge.
Federated	**I.** Heterogeneous pool of actors, that need to be identified **II.** Hierarchically integrated control structure, within and beyond company boundaries	There must be a clearly identifiable control structure, which controls resources as well as the outcome. Actors within the network come from different knowledge communities; knowledge is heterogeneous.
Anarchic	**I.** Heterogeneous & dynamic pool of actors / tools, that need to be dynamically identified & mobilized **II.** No centralized hierarchical control	No identifiable control structure. Outcome control is not existent. Actors within the network come from different knowledge communities; highly heterogeneous knowledge.

For the process of coding and displaying the results, we used MaxQDA Plus 12 and built upon on the recommendations of Mayring and Fenzl [27] to organize the statements made in the interview within the established framework.

In the next step, we carefully went through each established category and searched for emerging subcategories, which would point us towards trends within each type of innovation network. This helped distinguish between different forms of, for example, project innovation networks, thus, further structuring the results. An example of the process is depicted in figure 2:

Figure 2. Exemplary Process of Subcategory Coding



4 Results

In the following section, we present the results of the deductive qualitative analysis. During the analysis, a general, overarching theme emerged that underscored the importance of better understanding how pervasive digitization changes the way organizations cooperate and innovate. As IP 12 put it: "[Digitization] is already changing the world and is changing our world […]. Until now, we relied on creating value ourselves. This will not stay the case with digitization because we already believe that we cannot do it alone. But the question is: "Where do you differentiate yourself, where do you work together and where do you have the differentiation potential in the future?"

In the following subsections, we put the subcategories that emerged during the coding process in bold font at the beginning of each paragraph.

4.1 Project Innovation Network.

Overall, 16 out of 27 interview partners (IP) mentioned forms of cooperation that fit the established characteristics of project innovation networks. In the following, we list the subcategories we identified during analysis as depicted in figure 2.

Intra-departmental co-operation. Surprisingly, only 2 out of 16 interviewees (IP23, IP25) mentioned pure forms of project innovation networks in which all actors come from a specialized subdivision within the same company. It is characterized by a very homogeneous group of actors, which are working and specializing within one department, while avoiding external input. Furthermore, within the setting of an organization there is a clearly identifiable, high level of hierarchical control over resources and the expected outcome. For example, IP 25 stated: "We have a very strong silo or functional mindset - a strong optimization in individual areas, instead of making an end-to-end optimization."

Inter-departmental co-operation. Most interviewees (12 out of 16) pointed towards less clear-cut forms of project innovation networks as they involved inter-departmental cooperation. The networks still consist of actors from the same organization but cooperate beyond the boundaries of their respective department. Hence, the knowledge available in these inter-departmental settings is – while still being similar - less homogeneous than in the case of a silo mindset. IP06 described this trend by stating: "Departments used to be very well organized side by side – now everything is somehow dissolved a bit. It also produces much greater transparency. For better or worse." This trend of cooperating beyond the respective department boundaries manifests itself in the creation of idea platforms (IP02, IP26), increased proximity (IP03, IP04, IP12) and the formation of inter-departmental teams (IP01, IP02, IP05, IP11, IP16, IP20, IP21).

Idea platforms help generate new ideas and facilitate exchange between different departments and often take the form of corporate social media platforms. IP02 stated that: "We have a kind of corporate Facebook […]. You can post, like, upload media, share pictures, whatever. But there is also section where they [the company] can collect ideas […]. That means every employee can enter ideas". Furthermore, companies not only create virtual platforms to foster cooperation but also change the workspaces itself to increase the physical proximity between formerly separated departments. IP12 elaborated that: "We have created other premises. Especially, for initiatives that […] have to be processed quickly. We have an area here, […] to quickly establish cross-departmental relationships with colleagues in order to ideally prepare a product for the market in just a few weeks." The most common mechanism is the formation of inter-departmental project teams that pursue several goals such as entering new markets (IP01), quickly solving unexpected problems (IP11), strategy planning (IP16) and establishing new work processes (IP20). In order for such endeavors to be successful, numerous success factors were mentioned such as a good-will commitment of the executive board (IP01), the ability to focus 100% on the new task (IP01) and an open culture between different departments (IP05).

4.2 Federated Innovation Networks

During our conversations, federated innovation networks emerged as the most prevalent type of innovation network overall. Here, organizations strive to engage with highly heterogeneous knowledge (mainly) from outside the company and even from other markets while still maintaining a high level of control over resources and expected outcomes. In total, 26 out of 27 interviewees explicitly mentioned some or multiple forms of federated innovation networks. The most prominently mentioned ones are: Partnerships and cooperation (IP05, IP11, IP12, IP20-IP22, IP26), cooperation with startups and fintechs (IP02, IP05, IP08, IP10, IP20, IP25, IP26), customer panels (IP02, IP07, IP11, IP12, IP20, IP21), outsourcing (IP09, IP11, IP12, IP15), creation of platforms (IP01, IP11, IP15, IP24), external consultants (IP05, IP13, IP16, IP17), spin-offs (IP01, IP08, IP15). The most important ones, which were mentioned by at least five different interviewees, are explained in more detail below:

Partnerships and Co-operation were mentioned by 7 of the 26 different interviewees. As products and services are becoming increasingly complex [25], companies rely on an increasing number of co-operations to access external, heterogeneous knowledge and skills to solve problems that are out of their area of expertise. Control over the outcome and scope of such partnerships is typically ensured through contracts and agreements. IP20 stated: "I need other disciplines, other specialized [areas of] knowledge, which do not necessarily exist in our classical background of the company. In turn ... I said it before, I have to invest much more than I am used to in partnerships." Oftentimes companies collaborate with other companies that have expertise in fields such as the development of new IT solutions (IP13) or design choices (IP10, IP22). These co-operations aim at increasing customer value (IP11), extending research capabilities (IP20) and acquiring additional knowledge (IP26). The interviewees named an open-mindset on both sides (IP07), accepting failures in order to be able to try new things (IP21) and the development of strong networks (IP22) as important factors for success.

Startups and Fintechs. Even though co-operations with startups and fintechs could also be listed under the subcategory "partnerships and co-operations", 7 out of 26 interviewees specifically highlighted them as a special case of co-operation. startups were mentioned 6 times (IP02, IP05, IP08, IP10, IP25, IP27) and fintechs 3 times (IP02, IP10, IP20). The startup and fintech co-operations have the purpose of creating new ideas (IP05) and generally to increase value for customers (IP20, IP25). Furthermore, such co-operations are viewed as "a kind of research and development extension from outside [the company]. Of course, we only engage with ideas which we do not have to get out of our way." Furthermore, while they are perceived as an easy way to expand one's network and engage with new ideas, several interviewees pointed out common pitfalls such as that startup cooperation are great to learn but oftentimes do not end up profitable (IP05). Furthermore, startups and fintechs frequently compete for direct customer contact, which could transform incumbents into mere suppliers (IP20).

Customer Panels were mentioned by 6 out of 26 interviewees. Organizations try to closely align their offerings with what the customers want by directly acquiring feedback from their customers. To achieve this, companies create platforms on which customers can share suggestions for improvements or general feedback (IP03) or they even create an advisory board of specific customers to test products and quickly acquire feedback (IP03, IP07). The aim here is always to acquire customer feedback. However, it is crucial to build up the necessary processing capacity as it is easy to be overwhelmed by the amount of suggestions. IP03 stated for example: “The problem is, we have 1200 employees and 2 million customers. And 500 employees in the customer interface. They also get a lot of ideas from the customer and carry them there. So there are umpteen suggestions a day. They have to be evaluated. There must be someone who has time to look at his day-to-day business.”

4.3 Clan and Anarchic Innovation Networks

Interestingly, there were no statements that fulfilled the requirements for being classified as clan or anarchic innovation network. We will elaborate on this insight in the discussion section.

4.4 Mixed Forms of Project and Federated Innovation Networks

During the analysis, an additional mixed type of innovation network emerged. Such a mixed type of innovation network appears to help companies moving from one type of innovation network to another. Within this mixed innovation network type we distinguish between two subcategories: (1) transition mechanisms and (2) integration mechanisms.

Transition mechanisms are mechanisms that companies rely on to transition towards the north (more heterogeneous knowledge) or east (more distributed control). In total, we found 7 interviewees mentioning different examples such as: Creating new internal departments which become spin-offs (IP01, IP05), focusing on inter-departmental cooperation but involving some external actors (IP04, IP15, IP20, IP24, IP26) and creating internal startups (IP05). All of them are examples for the transition from project to federated innovation networks (i.e., north) or from project innovation networks towards a clan network (i.e., east).

An example for an east-transition is provided by IP01 who describes how the company went from full hierarchical control to a more distributed form of control by spinning off a project team and turning it into a more independent company: “It was a project team within Department 1 that […] has pushed this project forward […]. Until at some point – along with the market entry - the time was right to spin-off the project team into an independent sales GmbH. An independent business unit, which is one of our more successful business units to this day, because they have since made a very beautiful growth story.” Whereas it is questionable that this already qualifies as a full-fledged clan or anarchic innovation network, it certainly is a step towards this direction.

In general, the interviewees stated that such mixed forms are used to experiment with new business models or markets (IP01, IP05), quickly acquiring expertise from outside the company and speeding up the process of innovation (IP24) and for transformation projects (IP24).

Integration mechanisms are mechanisms that help integrate external actors, over which a company has little or no control. Thus, integration mechanisms help to move innovation networks towards the west (i.e., more centralized control) or south (i.e., more homogenous knowledge). Altogether, six interviewees mentioned integration mechanisms. Examples are acquisition (IP05), permanently integrating freelancers into work processes and decision-making (IP15) and classic hiring of valuable individuals (IP02, IP03, IP05, IP13). For example, IP 15 described how freelancers are completely integrated in decision making processes and feel as part of the team: "I also do not make the decision myself […], but I say: "The external and internal guys have decided on the design". There too, no separation. Freelancers already integrated internally. They also feel like internals. Connect with the company. Which is important - otherwise they cannot work the same as an internal one if they do not feel that it is their baby as well." Over time, the integrated freelancers develop a shared knowledge base with the company, thus, leading to a more homogenous knowledge base. Such a development signifies a slow move towards a project rather than a federated innovation network. Generally, these mechanisms were mentioned to help acquire new expertise (IP03, IP05, IP08, IP13) and to foster "out-of-the-box-thinking" (IP02).

5 Discussion

Theoretical Implications. This paper set out to answer two research questions. The first question is: *Which types of innovation networks are currently prevalent and why do organizations use them?* The main insight is that incumbent companies appear to rely only on project and federated innovation networks but not on clan and anarchic innovation networks, thereby, always maintaining control over resources and the eventual outcome. Furthermore, we found that organizations use project and federated innovation networks for different purposes. Project innovation networks are mainly used to facilitate idea generation, and to allow for fast problem solving and experimentation. Federated innovation networks on the other side are mainly used to gain access to external expertise, and to increase customer knowledge and thus customer value. Investigating which types of innovation networks are currently prevalent and why organizations use them addresses the "need to examine to what extent organizations simultaneously engage in multiple different types of networks, and how the intensity and proportion of these engagements affects the level and nature of their innovation work" [1].

The second question is: *How can companies fluidly transition between different types of networks?* We found that companies gradually transition between project and federated innovation networks by using mixed forms of innovation networks, which we categorized as either transition or integration mechanism. These identified mixed forms help address the fact that the two dimensions of control and heterogeneity are not as clear-cut in reality as they are theoretically [1]. By recognizing and categorizing mixed forms between different types of innovation network, we are able to take a more granular look and position the identified forms of innovation networks more precisely within the framework. Thereby, contributing to extant literature.

In the following, we discuss the highlights of our results and point out avenues for future research. When we started to analyze the data, we expected to find examples for distributed forms of control over the innovation network, the actors and the eventual outcome. This would be in line with the prediction of Lyytinen et al. [1] which stated that due to the ever more prevalent digitization there would be a trend towards more distributed forms of coordination and control. However, as stated above, our results do not support the prediction that there is a trend towards more distributed forms of control. For future research, it would be a worthwhile endeavor to investigate why organizations currently choose not to use clan and anarchic networks. A possible explanation is that while products become increasingly complex [8], [22], incumbents must engage their entire supply chain in order to successfully innovate (e.g., [29]). Thus, certain actors within a network of suppliers specialize for specific components which are then assembled into the final product by the manufacturer [24]. For such a modular design to be successful, manufacturers on top of a supply chain must maintain tight, centralized control over the innovation network and ensure that the individual components are modular and fit together [24]. Nonetheless, as Lyytinen et al. [1] stated, we can already observe examples of distributed control such as in open-source communities (e.g., [16], [30]). Projects such as Linux prove that the creation and development of highly complex products is feasible without any formal hierarchy [1], [24]. Thus, even though the incumbents in our case studies are currently not experimenting with more distributed forms of control, this may only be due to the tried-and-tested ways traditional companies always conducted their business. Path dependence, i.e. continuing to do things the traditional way, has implications for any renewal or restructuring process (e.g., [31]). Henceforth, after the successful emergence of open-source networks in highly digitized industries such as software and electronics, we expect "other industries [to follow] this trend as they embrace digitization in their products and services – for example, through implementing Internet of Things, digital product libraries, 3D printing and big data" [1, p.69]. Future research is needed to fully understand the factors that favor or hinder distributed forms of control and the decision of organizations to not use certain innovation networks.

Our results do support the prediction of Lyytinen et al. [1] that there is a strong trend towards more heterogeneous types of innovation networks. This is displayed in two different ways. First, 16 out of 27 interviewees mentioned project innovation networks. However, intra-departmental co-operation (co-operation within the same department; very homogenous setup) was only mentioned twice, whereas the overwhelming majority (12 of 16) referred to inter-departmental cooperation (co-operation across various departments; less homogenous as different fields of expertise within the company are involved). Thus, even within project innovation networks appears to be a strong trend towards the north. Second, multiple forms of federated innovation networks were mentioned by 26 of 27 interviewees, hence, providing a strong indication that the involvement of more heterogeneous knowledge is becoming increasingly important. This development provides many opportunities for further investigation as each form of cooperation has different advantages and disadvantages. Quantitative research might be able to further analyze these factors and relate them to innovation success or failure.

Limitations and Practical Implications. As for every research endeavor, it is important to understand and point out possible limitations. The chosen form of research design cannot claim to produce generalizable results but rather aims at conceptual clarity and identifying valuable areas of further inquiry. Furthermore, our sample of 27 interviews has a strong focus on the financial and manufacturing industry, which may results in biases. Furthermore, our case studies were conducted by exclusively interviewing high-level executives in incumbent companies. Thus, it would be interesting to contrast our findings with insights in upcoming, less established companies that might or might not rely on very different innovation networks including distributed forms of control. Hence, we welcome future studies to supplement our results with insights derived from qualitative or quantitative research.

The results of our research offer numerous implications for practitioners. First, our results indicate that different types of innovation networks are used for different purposes. Practitioners are well advised to focus on inter-departmental project innovation networks to facilitate idea generation, rapid experimenting and problem-solving if the knowledge necessary to address such issues already exists within the company. If there is the need to acquire additional expertise that does not exist within the company, practitioners can use several forms of federated innovation networks. Federated innovation networks were pointed out to be especially powerful to access external expertise. Furthermore, federated innovation networks can be utilized to directly get in touch with the customer and acquire real-time feedback, thus, providing ample opportunity to increase eventual customer value.

Second, even though there are successful examples of distributed control in some industries, the incumbents in our case studies appear to not actively experiment with distributed forms of control. Hence, here might lay new opportunities for innovative forms of collaboration and new business models.

Third, companies trying to slowly increase the heterogeneity of their resources and knowledge without disrupting their current processes can experiment with transition mechanisms. If there is a need to increase control and streamline heterogeneous knowledge resources into a more consistent, homogenous knowledge base practitioners can rely on integration mechanisms.

In conclusion, this paper produced new insights into how companies can move from one type of innovation network to another and which innovation networks are currently prevalent among incumbents. Most importantly, our results produce novel insights about why different types of innovation networks are used and point towards fruitful avenues for future research on innovation networks.

References

1. Lyytinen, K., Yoo, Y., Boland Jr., R.J.: Digital Product Innovation Within Four Classes of Innovation Networks. Info Systems J 26, 47–75 (2016)
2. Nambisan, S., Sawhney, M.S.: The Global Brain. Your Roadmap for Innovating Faster and Smarter in a Networked World. Wharton School Publ, Upper Saddle River, NJ (2007)
3. Borgatti, S.: The Network Paradigm in Organizational Research. A Review and Typology. Journal of Management 29, 991–1013 (2003)
4. Chesbrough, H.W.: Open Innovation. The New Imperative for Creating and Profiting From Technology. Harvard Business School Press, Cambridge, MA (2003)
5. Hippel, E.A. von, Katz, R.: Shifting Innovation to Users Via Toolkits. Management Science 49, 821–833 (2002)
6. Afuah, A., Tucci, C.L.: Crowdsourcing as a Solution to Distant Search. Academy of Management Review 37, 355–375 (2012)
7. Fichman, R.G.: Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods. Journal of the Association for Information Systems 5, 314–355 (2004)
8. Yoo, Y., Boland, R.J., Lyytinen, K., Majchrzak, A.: Organizing for Innovation in the Digitized World. Organization Science 23, 1398–1408 (2012)
9. Benner, M.J., Tushman, M.L.: Reflections on the 2013 Decade Award--"Exploitation, Exploration, and Process Management:. The Productivity Dilemma Revisited" Ten Years Later. Academy of Management Review 40, 497–514 (2015)
10. Majchrzak, A., Malhotra, A.: Towards an Information Systems Perspective and Research Agenda on Crowdsourcing for Innovation. The Journal of Strategic Information Systems 22, 257–268 (2013)
11. Saldanha, T.J.V., Mithas, S., Krishnan, M.S.: Leveraging Customer Involvement for Fueling Innovation. The Role of Relational and Analytical Information Processing Capabilities. MISQ 41, 367–396 (2017)
12. Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M.: Digital Innovation Management. Reinventing Innovation Management Research in a Digital World. MISQ 41, 223–238 (2017)
13. Drucker, P.F.: The Practice of Management. Harper & Brothers Publishers, New York, NY (1955)
14. Chesbrough, H.W.: Open Business Models. How to Thrive in the New Innovation Landscape. Harvard Business School Press, Boston, Mass. (2006)
15. Lakhani, K.R., Panetta, J.A.: The Principles of Distributed Innovation. Innovations: Technology, Governance, Globalization 2, 97–112 (2007)
16. Hippel, E.v.: Democratizing Innovation. MIT Press, Cambridge Mass. u.a. (2006)
17. Boudreau, K.J., Lakhani, K.R.: Using the Crowd as an Innovation Partner. Harvard Business Review 91, 60–69 (2013)
18. Hagedoorn, J., Duysters, G.: The Effect of Mergers and Acquisitions on the Technological Performance of Companies in a High-tech Environment. Technology Analysis & Strategic Management 14, 67–85 (2002)
19. Hildebrandt, B., Hanelt, A., Firk, S., Kolbe, L.M.: Entering the Digital Era – The Impact of Digital Technology. Related M&As on Business Model Innovations of Automobile OEMs. In Proceedings of the 36th International Conference on Information Systems (2015)
20. Mentzas, G., Apostolou, D., Kafentzis, K., Georgolios, P.: Inter-Organizational Networks for Knowledge Sharing and Trading. Inf Technol Manage 7, 259–276 (2006)
21. Tushman, M.L., Nadler, D.A.: Information Processing as an Integrating Concept in Organizational Design. Academy of Management Review 3, 613–624 (1978)

22. Yoo, Y., Henfridsson, O., Lyytinen, K.: The New Organizing Logic of Digital Innovation. An Agenda for Information Systems Research. Information Systems Research 21, 724–735 (2010)
23. Lusch, R.F., Nambisan, S.: Service Innovation. A Service-Dominant Logic Perspective. MISQ 39, 155–175 (2015)
24. Lee, J., Berente, N.: Digital Innovation and the Division of Innovative Labor. Digital Controls in the Automotive Industry. Organization Science 23, 1428–1447 (2012)
25. Boland, R.J., Lyytinen, K., Yoo, Y.: Wakes of Innovation in Project Networks. The Case of Digital 3-D Representations in Architecture, Engineering, and Construction. Organization Science 18, 631–647 (2007)
26. Yin, R.K.: Case Study Research and Applications. Design and Methods. SAGE, Los Angeles, London, New Dehli, Singapore, Washington DC, Melbourne (2018)
27. Mayring, P., Fenzl, T.: Qualitative Inhaltsanalyse. In: Baur, N., Blasius, J. (eds.) Handbuch Methoden der empirischen Sozialforschung, pp. 543–556. Springer Fachmedien Wiesbaden, Wiesbaden (2014)
28. Creswell, J.W., Creswell, J.D.: Research Design. Qualitative, Quantitative, and Mixed Methods Approaches. SAGE, Los Angeles (2018)
29. Powell, W.W., Koput, K.W., Smith-Doerr, L.: Interorganizational Collaboration and the Locus of Innovation. Networks of Learning in Biotechnology. Administrative Science Quarterly 41, 116–145 (1996)
30. Fitzgerald: The Transformation of Open Source Software. MISQ 30, 587–598 (2006)
31. Karim, S., Kaul, A.: Structural Recombination and Innovation. Unlocking Intraorganizational Knowledge Synergy Through Structural Change. Organization Science 26, 439–455 (2015)

Characterising Social Reading Platforms—
A Taxonomy-Based Approach to Structure the Field

Kristin Kutzner[1], Kristina Petzold[2], and Ralf Knackstedt[1]

[1] University of Hildesheim, Institute for Economics and Information Systems, Hildesheim, Germany
{kutznerk, ralf.knackstedt}@uni-hildesheim.de
[2] University of Hildesheim, Institute for Creative Writing and Literary Studies, Hildesheim, Germany
{petzold}@uni-hildesheim.de

Abstract. Due to digitalisation the way of discussing on books changes. Especially social reading platforms foster a more sociable and collaborative mentality. Manifold platforms exist that differ, for instance by enabling different formats to discuss on books, such as book club, forum or direct text work. As no consolidated overview of social reading platforms exists, this study aims at deriving a taxonomy of social reading platforms to structure the field. Based on (1) an initial literature review and (2) a collection of social reading platforms, we derived (3) the taxonomy that considers essential characteristics. Further, through (4) a cluster analysis, four types of platforms could be identified. Overall, based on the taxonomy researchers and practitioners can be informed within the variety of platforms, to compare, refine and develop further social reading platforms.

Keywords: Digital Transformation, Creative Industries, Digital Humanities, Social Reading Platforms, Taxonomy.

1 Introduction

"*The digitalisation is likely to change written culture, and therefore, at the same time society, even more than the development of Gutenberg's book printing*."
(translation of Volker Oppmann, CEO of the social reading platform Mojoreads)

Digitalisation refers to changes in various fields and influences human life [1]. Due to digital transformation, the way of communication, collaboration and participation within the entire society is changing [2, 3]. Creative industries (e.g., art, music, film) including book industry, as Volker Oppmann already stated, are highly affected by digital transformation as new information technologies enable changed ways of discussion on cultural and aesthetic practices. A variety of social reading platforms—the wide range will still increase—fosters a more sociable and collaborative mentality, enabling different formats to discuss and share opinions on books in the community. Consequently, digitalisation will lead to changed reading attitudes, habits and practices, resulting in novel realities and discourses [4]. Further, a notable part of German book sales—about 18.8% of sales of the German book trade in 2017 [5]—takes part on the

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Internet. According to [5], the estimated sales of Internet book sellers in 2004 accounted for 506 Mio. €, and in 2017 for 1.710 Mio. €, which is an increase of about 3.4 times over 13 years. Especially social reading platforms become more important for sellers and publishers, as they offer new opportunities, to improve the visibility of books for sales and book recommendations [4] and to address trends, influencing future product developments. For instance, LovelyBooks records more than 1.5 Mio. users per month and 70% of the users regularly buy books which they discovered on LovelyBooks [6].

Although social reading platforms are of growing interest, these platforms are very heterogeneous. That is why it is difficult to maintain an overview of their characteristics. Prior studies provided an overview of existing social reading platforms [4, 7] or a first proposal to structure social reading spaces [8, 9], focussing on limited characteristics of social reading. Nevertheless, as there is a need for "a taxonomy to make sense of a range of behaviours all which fit within the current 'social reading' rubric" [8], to the best of our knowledge, no consolidated overview exists. In order to promote this evolving field, we aim to develop a taxonomy that structures the field of social reading. Accordingly, this study aims to answer the following research question:

What kind of characteristics are provided by social reading platforms and how can these characteristics be structured?

To answer this question, we aim to derive a taxonomy of social reading platforms. A taxonomy facilitates the characterisation and analysis of a certain domain and supports ordering the complexity and provides a foundation for IS research, potentially leading to future research directions [10]. Both for research and practice, the taxonomy can be used to derive implications for future social reading platforms. Researchers are, for example, supported to position and redesign their platforms or research results, and to identify gaps that need to be addressed by future endeavour. Practitioners (book sellers, publishers), for instance, can be informed within the wide range of social reading platforms, to compare, refine and develop further platforms. In addition, users (e.g., reviewers, authors) are supported to make an informed decision concerning the selection of a social reading platform as they receive an overview of platforms and their characteristics. For developing the taxonomy, we first outline the background of social reading and social reading platforms (Section 2). Following our research design (Section 3), we gathered initial literature, collected social reading platforms and applied a taxonomy-development approach to classify the platforms based on specified characteristics. Further, we identified types of social reading platforms by carrying out a cluster analysis (Section 4). Afterwards we discuss the limitations, results and implications (Section 5) and conclude with our main findings (Section 6).

2 Background

In this section, we introduce the concepts of social reading, social reading platform and attempts in research to structure social reading platforms.

Social reading platform. Traditionally, reading is associated to be done alone, underlining text passages and writing notes in the margins. However, readers have always communicated their impressions and evaluations about what they read, restricted to small communities (e.g., traditional book clubs) [4]. Thus, reading itself as it functions as a vehicle of human interaction by transferring thoughts and emotions between people can be defined as a social activity. According to various researchers, digitalisation changes the basic structures of communication before, during and after reading [11] and enables a changed form of socialisation of reading that encourages collaboration and exchange, resulting in a more and more social reading [4, 12, 13]. In addition, social reading can be described as a form of collective reading of digital texts and communication in reader-centred communities in digital networks [11]. An Internet platform or a specific software product is the meeting place for social reading in the above described sense. Taking part in such an application, users can communicate, exchange information and share opinions on books [14]. To do so, for instance, the platforms allow users to write reviews on books or participate in books clubs.

Structuring social reading platforms. Due to the increasing relevance of social reading platforms, some studies already attempt to structure and specify them. Nonetheless, by conducting a literature review[1], we found only a few articles that provide a first structure. Whereas Cordón-García et al. [4] and Boesken [7] provided an overview of existing social reading platforms, Stein [8] provided a taxonomy of social reading in order to make sense of behaviours that fit with the current social reading practices. However, he focuses on technology and considers online discussion used in social reading in a more general way. Moreover, Winget [9] provides an augmented taxonomy of social reading, reflecting background in commonplace books and annotation behaviour. Following, to the best of our knowledge, there is no research providing a consolidated structure of social reading platforms, and there still seems to be a need for a widely accepted overview. Accordingly, in this study we aim to close these gaps by providing a proper basis for further research in the field of social reading.

3 Research Design

In order to develop a taxonomy of social reading platforms, and therefore, to address our research question, we conducted a four-staged research design. As first steps, we carried out (Stage 1) an initial literature review and (Stage 2) collected social reading platforms. Following, based on the findings, we (Stage 3) iteratively built a taxonomy to structure the platforms obtained and performed (Stage 4) a cluster analysis to identify types of social reading platforms (Figure 1).

Stage 1: Initial literature review. As a first step for characterising social reading platforms, we conducted a literature review based on the rigorous procedure of vom Brocke et al. [15]. In March and April 2018, we searched for relevant literature in the scientific domain (that is book chapters, books, conference papers, working papers and journal publications) in both IS and literary studies. Based on our definition of social reading (Section 2), we used the following combinations of keywords in AISeL, Google

[1] We searched for existing research that structures social reading platforms. We used "Social Reading Platform" AND ("Classification" OR "Taxonomy") as search phrase in AISeL and Google Scholar (accessed on July 2018).

Scholar, BDSL and university library catalogue: *book, literature, social reading, platform, forum, community, book seller, online shop*. After reading title, abstract and keywords, we finally considered seven scientific sources (two books, three book chapters, one journal publication, and one working paper) to define potentially initial dimensions and characteristics of the taxonomy.

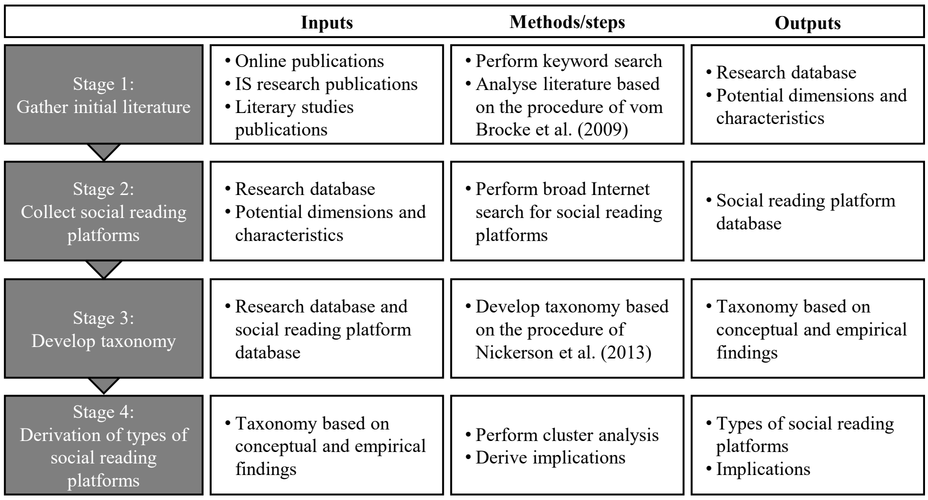


Figure 1. Research design

Stage 2: Collection of social reading platforms. Based on the results from the initial literature review and the researchers' expertise in the area of social reading platforms, in May and June 2018, we investigated a number of social readings platforms combined with a broad Internet based search for social reading platforms. To contribute to comparability, we included German platforms as Germany is one of the largest European markets in terms of publishers' turnover [16]. We excluded platforms that do not longer exist—although they are still listed in the Internet—or have less emphasis in the community (measured by small user numbers). As a result, we obtained ten social reading platforms.

Stage 3: Taxonomy development. The development of taxonomies has often been used to structure and analyse fields both in IS [17–19] as well as especially in creative industries [8, 20]. Moreover, taxonomies can be seen as a step towards developing analytic theories [21, 22]. In our study, we build a taxonomy in line with the rigorous procedure of Nickerson et al. [10]. Following, we define *meta-characteristics* as the components of social reading platforms. Moreover, Collaboration Virtualization Theory [23] provides significant factors (e.g., task, technology, team) which influence the success of collaboration technology. In line, as it provides potential factors that might affect the usage of social reading platforms, we adopted the framework and extended the meta-characteristics. Further, we adopted the *objective and subjective ending conditions* proposed by Nickerson et al. [10]. According to Nickerson et al. [10], the taxonomy development process is an iterative one within we may choose between *empirical-to-conceptual* and *conceptual-to-empirical approach*. In sum, we run through 13 iterations, whereby after six iterations no more changes occurred (Figure 2).

As a *1st iteration* (*conceptual-to-empirical*), we integrated the taxonomy dimensions and characteristics identified during the literature review (Stage 1). As an initial result, we introduced ten dimensions: *type of communication* [4, 7, 8, 11, 24], *realisation of subsequent communication* [7, 11, 24, 25], *off-topic communication* [7], *system-based roles* [7, 24], *identity customisation* [7, 11, 24, 26], *provided link to other communities* [4, 7, 11, 24], *registration for the platform* [7], *visibility of content* [7, 11, 25], *financing of the platform* [7, 11] and *reviewer gratification* [11, 24, 26]. Based on these conceptual findings, in the *second, third, fourth, fifth and sixth iteration* (*empirical-to-conceptual*), the social reading platforms that we collected were investigated and classified by two researchers independently to contribute to the robustness.[2] In a subsequent workshop (*iteration 7*) the results were consolidated and structured within the taxonomy, resulting in (i) new dimensions: *type of cultural artefact addressed by subsequent communication*, *assistance for subsequent communication*, *social media intelligence*, *transaction offerings*, *author gratification* as well as (ii) additional characteristics. After the sixth iteration the taxonomy led to the successful classification of all social reading platforms from the database (detailed descriptions see Section 4). In contrast to Nickerson et al. [10], the characteristics of each dimension are not mutually exclusive.

Stage 4: Derivation of types of social reading platforms. Finally, to identify types of social reading platforms we performed a cluster analysis. This type of analysis is a widespread used analytical tool in IS research for grouping objects and investigating correlations in differently-sized samples from small datasets (e.g., 15 items) to larger ones (e.g., 9025 items). Especially, for further analysis of derived taxonomies, a cluster analysis is often used in IS research [27]. In line, we used the selected social reading platforms as items and the set of characteristics on the basis of the taxonomy as clustering variables. Following, we applied the K-means algorithm that is one of the most common methods for such a clustering [28]. Following [18, 29], we applied a two-step approach to identify clusters of social reading platforms from our taxonomy by using python module *scikit-learn* that provides a wide range of machine learning algorithms, both unsupervised and supervised [30].

1. Step (Ward's method). The ward's method is a procedure to build hierarchical clusters of subsets on the basis of their similarity. The method combines two closest subsets into one cluster and repeats this procedure until all subsets are in one cluster [31]. The amount of identical characteristics along the taxonomy determined the similarity between two subsets [18]. To follow the sequence in which the subsets have

[2] After the initial literature review (1) for each iteration, we considered a new social reading platform that we identified and performed subsequent workshops in order to consolidate results: (2) LovelyBooks (www.lovelybooks.de), (3) BuecherTreff (www.buechertreff.de), (4) Amazon (www.amazon.de), (5) Lectory (www.lectory.io), (6) Mojoreads (www.mojoreads.com), (7) Workshop, (8) WasLiestDu? (www.wasliestdu.de), (9) Buechereule (www.buechereule.de), (10) Literaturforum (www.literaturforum.de), (11) Buecher.de (www.buecher.de), (12) Hugendubel (www.hugendubel.de), (13) Workshop. After the sixth iteration, no more changes occurred.

been united in relation to the distances, we plotted a dendrogram. Regarding the significant jumps in the distance of the joint clusters, we identified three or four clusters as useful numbers. To highlight more different cluster, we chose four.

2. Step (K-means method). Second, we applied K-means method that divides data into clusters, minimising the within-cluster sum of squares [32]. We chose initial cluster centres using 'k-means++' and the algorithm iterated 300 times. Within each iteration, the algorithm run with ten different centroid seeds to get the best results.

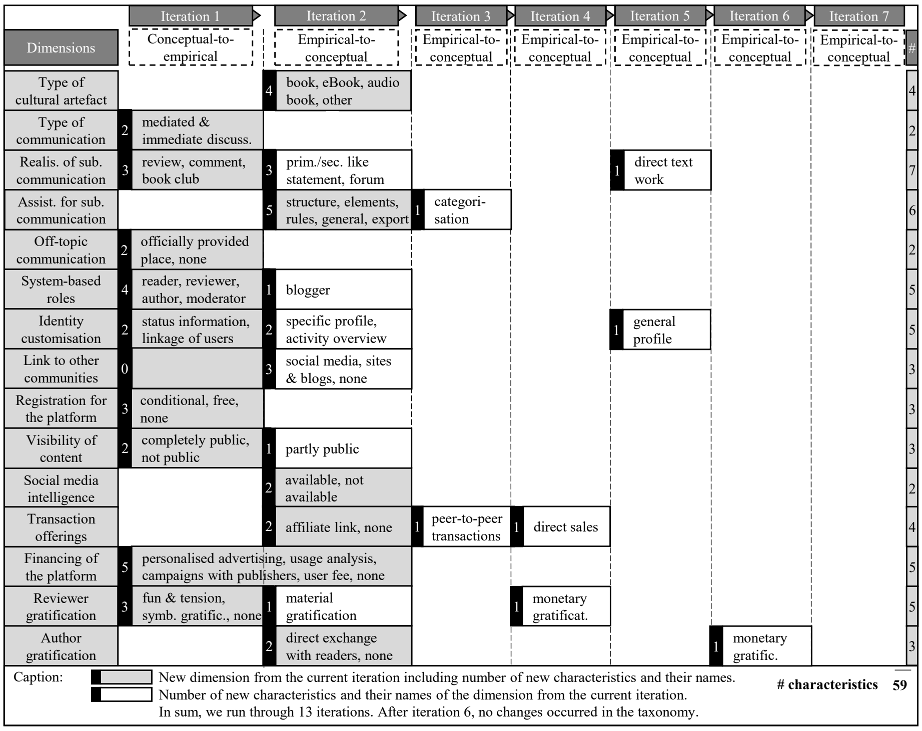


Figure 2. Development of dimensions and characteristics for the taxonomy

4 Characteristics of Social Reading Platforms

In this section, we present our main contribution, a taxonomy of social reading platforms and the derived types of social reading platforms.

4.1 A Proposed Taxonomy of Social Reading Platforms

The resulting taxonomy contains 15 dimensions, each with two to seven distinct characteristics (Figure 3). Each dimension and their corresponding characteristics are explained in the following. It is important to note that the taxonomy contains the most important dimensions along which the platforms differ. Consequently, components that are identical for all platforms (e.g., analysis of personal data, increasing publicity of books as gratification) are not listed here. Further, each dimension relate to one or more categories of Collaboration Virtualization Theory *task, community, technology* [23] that we extended with *business* as the taxonomy considers business issues as well.

Dimensions	Characteristics							
Type of cultural artefact addressed by subsequent communication (**TA**, B)	book	eBook	audio book	other				
Type of communication (**TA**, C, TE)	mediated discussion on a book (change of medium)	immediate discussion on a book (in the margins)						
Realisation of subsequent communication (**TA**, C, TE)	primary thematisation Writing a review	secondary thematisation Comment function	primary thematisation Like statement	secondary thematisation Like statement	book club: one type / diff. types	discussion forum	direct text work (visual marks)	
Assistance for subsequent communication (**TA**, C, TE)	prescribed structure of a review	proposed elements of a review	direct instructions	general instructions	writing rules & tone	general assistance	categorisation of reviews	export function
Off-topic communication (**TA**, C, TE)	officially provided place	none						
System-based roles (TA, **C**, TE)	reader	reviewer	author	moderator	blogger			
Identity customisation (TA, **C**, TE)	general user profile	specific user profile	status information	activity overview	linkage of users			
Provided link to other communities (**C**, TE)	social media	personal sites & blogs	none					
Registration for the platform (**TE**, B)	conditional registration	free registration	none					
Visibility of content (**TE**)	completely public	partly public	not public					
Social media intelligence (**TE**, B)	available	not available						
Transaction offerings (C, TE, **B**)	direct sales	affiliate link	peer-to-peer transactions	none				
Financing of the platform (**B**)	personalised advertising	usage analysis	campaigns with publishers	user fee	none			
Reviewer gratification (TA, C, **B**)	monetary gratification	material gratification	symbolic gratification	fun & tension	none			
Author gratification (TA, C, **B**)	direct exchange with readers	monetary gratification	none					
Caption:	Categories: TA = task, C = community, TE = technology, B = business	A dimension can be assigned to different categories. However, each dimension can be assigned to a most relevant category that is presented with bold letters.						

Figure 3. Taxonomy of social reading platforms

Type of cultural artefact addressed by subsequent communication. Using social reading platforms to discuss and share opinions within a community, not only *books* but also *eBooks*, *audio books* or *other* cultural artefacts (e.g., public reading by authors, music or films) can be focused.

Type of communication. First, platforms allow people to write and share a review —separated from the book text itself—and to engage in subsequent discussions on specific books (*mediated discussion on a book—change of medium*). Second, providing a dynamic margin on the page of a book, allows a close discussion on books. The composition of multiple comments forms a review (*immediate discussion on a book—in the margins*) [8].

Realisation of subsequent communication. Social interaction related to social reading can be shaped by the theoretical concept of subsequent communication. It defines communication that is directly attached to the book [7, 11]. People directly write a review on a book (*primary thematisation—writing a review*) or they comment on reviews of other people (*secondary thematisation—comment function*). Further, numerical ratings (e.g., star, heart or thumb ratings) both for the book in general (*primary thematisation—like statement*) or for reviews or comments of other people (*secondary thematisation—like statement*) can be made. In addition, *book clubs* allow the structured discussion on books, offering different types of application (e.g., discussion on a chapter is only possible/open, after everyone read the chapter or at any time, readers are allowed to write comments to each chapter). For instance, authors or publishers can initiate a book club for a selected book. Interested people can participate—sometimes restricted by application procedures or number of participants—after having read the book. Thus, the initiator of the book club can start the discussion on the book, providing a structured selection of themes (e.g., by chapters of a book) that have to be discussed, resulting in multiple reviews and comments. The initiators get customised, structured feedback and the participants can engage in clear discussions, to communicate their feelings and evaluations about what they read. Moreover, people are supported to address additional literary themes, using a *discussion forum* and sometimes *direct text work (visual marks)* on books (e.g., use of text marking, bookmarks) is possible.

Assistance for subsequent communication. While writing a review, some platforms *directly* provide a *prescribed structure of a review* (e.g., reviews' title, short summary, positive and negative issues, resume) and/or *directly propose elements* that might be *discussed* (e.g., writing style, story of the book, figures). In contrast, both for *structure* and *elements general instructions* that are not simultaneously provided while writing a review are possible (e.g., first entry of a discussion forum that explicates how a review can be structured). In addition, *writing rules and tone* (e.g., rules of courtesy as avoiding swear words and personal insults, observance of spelling), *general assistance* for dealing with the platform and its characteristics (e.g., FAQ document, an introductory learning unit after registration) and a *categorisation of reviews* (e.g., verified sale) are provided. Moreover, some platforms allow *exporting* the text of subsequent communication.

Off-topic communication. Further, off-topic communication (i.e., communication without direct relation to literary themes) [7] can be performed on *officially provided places*. For example, some platforms provide an independent discussion forum or support exchange of messages between people (e.g., via chat or pin board).

System-based roles. To receive content on a platform and to participate in discussions on books, several system-based roles are established [7]. People can read content (e.g., reviews, discussions in book clubs) (*reader*) and participate in discussions as *reviewers*, mostly after registration. Further, people can register as *author*, *blogger* or *moderator* with special rights (e.g., generation und control of book clubs).

Identity customisation. People can create a virtual identity, to be noticeable for others in order to get an idea of each other, and to build an important foundation for trust [7]. Their virtual identity can be customised via an user profile (*general user or specific user profile*), several forms of *status information* (e.g., number of received "likes", rank measured by number of reviews and their helpfulness assessed by others), *activity overview* (e.g., reading statistics, wish list of books) and *linkage to other users* (e.g., "following" users, finding "friends").

Provided link to other communities. Further, the connection to other communities, offering links to *social media* (e.g., connection with Facebook account, social media button) or *personal sites and blogs* (e.g., link to book blog) can be provided.

Registration for the platform. To take part in discussions and to make use of all functions, a *free registration* can be required. Sometimes, a platform might be a more "closed system", providing access to a selected number of people (e.g., after invitation) (*conditional registration*).

Visibility of content. The general public is able to read, and therefore, to receive the whole content, presented on a platform (*completely public*). However, sometimes the visibility of content might be partly restricted to public (*partly public*) (e.g., author profiles are visible, whereas visibility of reviewer-profiles depends on registration) or only visible for a closed community (*not public*).

Social media intelligence. Aiming to collect, monitor, analyse, summarise and visualise data of the platforms, social media intelligence is integrated (*available*) [33]. For instance, a similarity function provides similar users within the community that are sometimes called "book neighbours". Further, similar authors or books are recommended that support the selection of further books.

Transaction offerings. Platforms directly support purchasing a book (*direct sales*) or guide people by an *affiliate link* to an external online-shop that, in case of purchase decision, financially supports the platform. Besides, *peer-to-peer transactions* allow exchange of cultural artefacts (e.g., "buy, exchange, gift forum").

Financing of the platform. To finance a platform, several activities are addressed by providers. *Personalised advertising* (e.g., newsletter, press advertising) or *usage analysis* to contribute to market and opinion research and to improve services are sample activities. To do so, external services are frequently used (e.g., Google Analytics). Further, *campaigns with publishers* (e.g., book club initiated and controlled by publishers, free copies) or *user fees* are possible financial sources.

Reviewer gratification. Next to more social gratifications of society (e.g., presentation and confirmation of peoples' own reading quality, belonging to a community) [11], some reviewer gratification can be derived from the platforms. People can have *fun and tension*, participating in specific games (e.g., writing competition, riddles), they can gain *symbolic,* (e.g., status enhancement because of users' activities), *monetary* or *material gratification* (e.g., draw of books).

Author gratification. Besides, authors benefit as they *directly interact with readers*, receiving immediate feedback. Also, *monetary gratifications* are possible.

4.2 Types of Social Reading Platforms

Following the cluster analysis method (Section 3), we identified four clusters. Each cluster has a different focal point along the dimensions and characteristics of social reading platforms. As the characteristics within a dimension are collectively exhaustive, the results can be read as percentages. Moreover, the absolute frequency of each characteristic is presented. For example, 100% of the social reading platforms in Cluster

1 support mediated discussion on a book as type of communication. In sum, the majority of platforms (9/10) supports such a mediated discussion (Figure 4). The darker the colour of a cell, and therefore, the higher the percentage of a characteristic within a dimension, the more it is shaping a cluster. Next, we present the clusters, highlighting the most typical characteristics of each cluster and utilising demonstrative examples.

Cluster 1—manifold discussions within a bonded community. The *mediated discussion on books* has been changed by digitalisation. Providing manifold *realisations of subsequent communication*, people interested in literature—especially *reviewers and authors*—can engage in meditated discussions, to communicate their feelings and evaluations about what they read. For instance, book clubs that allow a structured discussion on books, enable authors a *direct exchange with readers*, resulting in immediate feedback. To encourage collaboration and exchange, platforms provide various *assistance for subsequent communication*. Further, discussions on books are directly related to a bonded community. Therefore, users are closely linked, providing *customised identities* and expanding communication by non-literary themes (*off-topic communication*). In line, people benefit from specific games related to having *fun and tension*, from *material* (e.g., draw of books) or *symbolic gratifications* (e.g., on LovelyBooks users get "dog-ears" for their activities on the platform). Further, *affiliate links* and *peer-to-peer transactions* are provided. To reach many people, a lot of content is *visible for public* and only *free registration* in order to participate is required. For financing a platform, *personalised advertising*, *usage analysis* and *campaigns with publishers* are used.

Cluster 2—assessment of books to support purchase decisions. Offering *direct sales* transactions, platforms allow people to write reviews and numerically rate several (e.g., star ratings) *types of cultural artefacts*. Although secondary thematisation is allowed, manifold discussions between people—as they are observed in Cluster 1—are not focused (e.g., no official interaction between readers and authors). In line, less *assistance for subsequent communication* is provided and identity customisation is restricted to the users' own characteristics (*specific user profile*, *status information*, *activity overview*) without special interest in linkage of users (e.g., *no off-topic communication*). Besides, the visibility of content is *completely public* and reviewer can benefit from *symbolic*, *material* or *monetary gratifications* (e.g., users win a book voucher after writing a review on buecher.de). Moreover, *personalised advertising* and *usage analysis* are used to finance the platform.

Cluster 3—immediate discussions on books within a closed community. Providing a dynamic margin on the page of an *eBook*, allows *immediate discussions on books*. Only after invitation of *moderators*, *reviewers* can register (*conditional registration*) and participate. Consequently, the content is *invisible for public*. Engaging in discussions, *book clubs*, private *discussion* groups and *direct text work (visual marks)* can be used. To do so, several *assistance for subsequent communication* is provided (e.g., prescribed structure and proposed elements). Besides, next to *personalised advertising*, *usage analysis* and *campaigns with publishers*, also *user fees* are introduced as financial sources.

Dim	Type of cultural artefact				Type of commun.		Realisation of subsequent communication							Link to other communities		
Charactaeristics	book	eBook	audio book	other	mediated discussion	immediate discussion	prim. them. writing a review	second. them. comment fct.	prim. them. like statement	second. them. like statement	book club	discussion forum	direct text work (visual marks)	social media	personal sites & blogs	none
#*	9	7	7	5	9	2	9	9	7	9	7	8	2	9	3	1
Cluster 1	35%	18%	29%	18%	100%	0%	19%	19%	13%	16%	16%	19%	0%	75%	25%	0%
Cluster 2	25%	25%	25%	25%	100%	0%	29%	14%	29%	29%	0%	0%	0%	50%	0%	50%
Cluster 3	0%	100%	0%	0%	0%	100%	0%	20%	0%	20%	20%	20%	20%	100%	0%	0%
Cluster 4	50%	50%	0%	0%	50%	50%	14%	14%	14%	14%	14%	14%	14%	50%	50%	0%

Dim	Assistance for subsequent communication						Off-topic commun.		System-based roles					Registration for platform		
Charactaeristics	prescribed struc. of a review	proposed elem. of a review	writing rules & tone	general assistance	categorisation of reviews	export function	officially provided place	none	reader	reviewer	author	moderator	blogger	conditional registration	free registration	none
#*	7	7	8	9	2	2	5	5	9	10	6	3	2	1	9	0
Cluster 1	23%	23%	19%	23%	8%	4%	67%	33%	30%	30%	25%	10%	5%	0%	100%	0%
Cluster 2	0%	0%	33%	67%	0%	0%	0%	100%	50%	50%	0%	0%	0%	0%	100%	0%
Cluster 3	25%	25%	25%	0%	0%	25%	0%	100%	0%	50%	0%	50%	0%	100%	0%	0%
Cluster 4	0%	0%	50%	50%	0%	0%	100%	0%	25%	25%	25%	0%	25%	0%	100%	0%

Dim	Identity customisation					Visibility of the content			Social media intel.		Transaction offerings			
Charactaeristics	general user profile	specific user profile	status information	activity overview	linkage of users	completely public	partly public	not public	available	not available	direct sales	affiliate link	peer-to-peer transactions	none
#*	3	7	7	10	7	4	5	1	9	1	5	5	4	0
Cluster 1	4%	24%	24%	24%	24%	33%	67%	0%	100%	0%	10%	50%	40%	0%
Cluster 2	0%	25%	25%	50%	0%	100%	0%	0%	100%	0%	100%	0%	0%	0%
Cluster 3	50%	0%	0%	50%	0%	0%	0%	100%	0%	100%	100%	0%	0%	0%
Cluster 4	33%	0%	0%	33%	33%	0%	100%	0%	100%	0%	100%	0%	0%	0%

Dim	Financing of the platform					Reviewer gratification					Author gratification		
Charactaeristics	personalised advertising	usage analysis	campaigns with publishers	user fee	none	monetary gratifcation	material gratification	symbolic gratification	fun & tension	none	direct exchange with readers	monetary gratification	none
#*	9	9	5	1	0	3	6	7	5	1	7	1	3
Cluster 1	38%	38%	25%	0%	0%	6%	29%	35%	29%	0%	83%	0%	17%
Cluster 2	50%	50%	0%	0%	0%	33%	33%	33%	0%	0%	0%	0%	100%
Cluster 3	25%	25%	25%	25%	0%	0%	0%	0%	0%	100%	100%	0%	0%
Cluster 4	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%	50%	50%	0%

* absolute frequeny of each characteristic

Caption: the darker the colour of a cell, the higher the percentage within a dimension

Cluster 1: LovelyBooks, BuecherTreff, Amazon, WasLiestDu?, Buechereule, Literaturforum, *Cluster 2*: Buecher.de, Hugendubel, *Cluster 3*: Lectory, *Cluster 4*: Mojoreads

Figure 4. Results of the cluster analysis

Cluster 4—hybrid discussions on books, related to sales and monetary gratification. Both *mediated* and *immediate discussion on books and eBooks* can be performed by several roles (*reader*, *reviewer*, *author*, *blogger*). In consequence, multiple *realisations of subsequent communication* are available. To offer a diverse view on literary themes, and therefore, to contribute to various discussions, *personal* *literary sites* and *blogs* are linked. In addition, people can *directly buy* books. If the purchase decision has been influenced by a certain users' comment, she or he will receive *monetary gratification* (Mojoreads).

5 Limitations, Discussion and Future Directions

Before discussing selected issues and possible research directions, we acknowledge the limitations of this study. First, our investigation is limited to the selected German social reading platforms and the general validity of our sample cannot be guaranteed, although the developed taxonomy fulfils the ending conditions. However, as Germany is one of the largest European markets in terms of publishers' turnover [16], it provides a useful starting point for our research. Further, we would argue that our results provide a useful characterisation of platforms and the derived types already indicate major differences, underlining the heterogeneity between social reading platforms. Second, the identification of dimensions and characteristics is based on our own decisions and interpretations. To contribute to the robustness and the reliability, two researchers independently analysed the platforms and subsequently consolidated their results. Third, the number of types of social reading platforms is affected by the interpretation of the dendrogram. Although we followed established approaches (Ward's method, K-Means method), another number of clusters might have influence on the results.

Despite the mentioned limitations, we believe this study to be an important first step in investigating how the digitalisation changes the literary culture. For reasons of space limitations, we discuss selected issues and present possible research directions below.

Derive implications for design features of social reading platforms. Analysing social reading platforms, various characteristics and four types have been identified. Considering further platforms, not only non-German social reading platforms but also platforms of a wider field (e.g., related to scientific literature), may support verifying or extending our results. In addition, the taxonomy provides a foundation for other representations (e.g., derivation of a data model) that might function as a detailed basis for comparison of platforms and further developments (e.g., derivation of new characteristics that could not be found in the investigated platforms). Besides, based on our findings researchers are provided to derive design features for social reading platforms, contributing to the design of review systems [34] in general as well. For instance, various realisations of *assistance for subsequent communication* exist. Sometimes a guideline for writing reviews (structural hints, proposed elements) is separately provided on a platform, whereas also direct instructions while writing a review are provided. Hence, it would be interesting to analyse the influence of such different forms of assistance on reviews. Doing so, the content and structure of reviews should be analysed and compared to the proposed guidelines, leading to enhanced and novel proposals for assistance that might support the manifold, detailed discussion on books. In line, as online reviews influence purchase decisions [35, 36], and economic outcomes, also publishers, authors and book sellers can benefit.

Engage in social reading intelligence. Regarding the selected social reading platforms, almost all of them provided *social media intelligence* (9/10) to derive actionable information from social reading activities. For instance, similarity functions provide similar users within the community that are sometimes called "book neighbours". Further, the majority of platforms performs *usage analysis* (9/10) in case of market and opinion research and improvement of services. As the digitalisation will lead to novel reading realities and discourses [4], adopted or new tools and algorithms are required to analyse the changed form of socialisation of reading. "Social reading intelligence" presents great potential with important practical relevance. For instance, content that emerges within multiple realisations of subsequent communication could be analysed. Making use of this knowledge, for example, the impact of book clubs on sales—similar to effect of reviews on sales [37]—might be investigated.

6 Conclusion

In order to develop a taxonomy of social reading platforms, we carried out an initial literature review, collected social reading platforms, analysed the findings and built a taxonomy that *structures* various *characteristics that are provided by social reading platforms*. Performing a cluster analysis, we discovered four types of platforms: (I) manifold discussions within a bonded community, (II) assessment of books to support purchase decisions, (III) immediate discussions on books within a closed community, (IV) hybrid discussions on books, related to sales and monetary gratification. Based on the findings, we discussed selected issues and possible future research endeavours.

Overall, our findings contribute to the ongoing research in digital transformation in creative industries. Based on the taxonomy, both research and practice are supported to derive implications for future social reading platforms. Researchers are, for example, supported to position and redesign their platforms or research results, and to identify gaps that need to be addressed by future endeavour. Practitioners (book sellers, publishers), for instance, can be informed within the wide range of social reading platforms, to compare, refine and develop further platforms. In addition, users (e.g., readers, reviewers, authors) are supported to make an informed decision concerning the selection of a social reading platform as they receive an overview of platforms.

7 Acknowledgements

This research was conducted in the scope of the research project "Rez@Kultur" (01JKD1703), which is funded by the Bundesministerium für Bildung und Forschung (BMBF). We would like to thank them for their support.

References

1. Kutzner, K., Schoormann, T., Knackstedt, R.: Digital Transformation in Information Systems Research: A Taxonomy-based Approach to Structure the Field. In: Proceedings of the 26th European Conference on Information Systems, Portsmouth (2018)

2. O'Reilly, T.: What is the Web 2.0?, http://www.oreilly.com/pub/a//web2/archive/whatis-web-20.html (2005)
3. Wilms, K.L., Meske, C., Stieglitz, S., Decker, H., Fröhlich, L., Jendrosch, N., Schaulies, S., Vogl, R., Rudolph, D.: Digital Transformation in Higher Education–New Cohorts, New Requirements? In: Proceedings of the 23d Americas Conference on Information Systems, Boston (2017)
4. Cordón-García, J.-A., Alonso-Arévalo, J., Gómez-Díaz, R., Linder, D.: Social Reading Platforms, applications, clouds and tags. Chandos Publishing, Oxford (2013)
5. Börsenverein des Deutschen Buchhandels: Statista-Dossier zum Buchmarkt, https://de.statista.com/themen/1069/buchmarkt/ (2017)
6. LovelyBooks: LovelyBooks Mediadaten 2018, https://s3-eu-west-1.amazonaws.com/media.lovelybooks.de/LovelyBooks_2018.pdf (2018)
7. Boesken, G.: Literarisches Handeln im Internet Schreib- und Leseräume auf Literaturplattformen. UVK, Konstanz (2010)
8. Stein, B.: A Taxonomy of Social Reading: a proposal, http://www.futureofthebook.org/social-reading/matrix/ (2010)
9. Winget, M.: A Mediation on Social Reading and Its Implications for Preservation. Dig Tech Cult. 42(1), 1–14 (2013)
10. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. Eur. J. Inf. Syst. 22(3), 336–359 (2013)
11. Kuhn, A.: Lesen in digitalen Netzwerken. In: Rautenberg, U., Schneider U. (eds.) Lesen Ein interdisziplinäres Handbuch. pp. 427–444. De Gruyter, Berlin, Boston (2015)
12. Cavallo, G., Chartier, R.: A History of Reading in the West. University of Massachusetts Press (2003)
13. Manguel, A.: A History of Reading. Penguin Books (1996)
14. Pouliot, C.: Lectura social y medios digitales: la revolución de los receptores en la Era 2.0. Congreso Amigos 2011: El usuario, la información y la biblioteca de los Ángeles, Mexico (2011)
15. vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on the importance of rigour in documenting literature search process. In: Proceedings of the 17th European Conference on Information Systems, Verona (2009)
16. Federation of European Publishers: European book publishing statistics, https://fep-fee.eu/European-Book-Publishing-922 (2017)
17. Haas, P., Blohm, I., Leimeister, J.M.: An empirical taxonomy of crowdfunding intermediaries. In: Proceedings of the 35th International Conference on Information Systems, Auckland (2014)
18. Remane, G., Nickerson, R., Hanelt, A., Tesch, J.F., Kolbe, L.M.: A taxonomy of carsharing business models. In: Proceedings of the 37th International Conference on Information Systems, Dublin (2016)
19. Szopinski, D., Schoormann, T., John, T., Knackstedt, R., Kundisch, D.: How software can support innovating business models: A taxonomy of functions of business model development tools. In: Proceedings of the 23d Americas Conference on Information Systems, Boston (2017)
20. Katsma, C., Spil, T.: A taxonomy of digital music services. In: Proceedings of the 16th Americas Conference on Information Systems, Lima (2010)
21. Gregor, S.: The Nature of Theory in Information Systems. MIS Q. 30(3), 611–642 (2006)
22. Williams, K., Chatterjee, S., Rossi, M.: Design of emerging digital services: a taxonomy. Eur. J. Inf. Syst. 17(5), 505–517 (2008)

23. Fan, S., Sia, C.L., Zhao, J.L.: Towards Collaboration Virtualization Theory. In: Proceedings of the 16th Pacific Asia Conference on Information Systems, Ho Chi Minh City (2012)
24. Lukoschek, K.: Ich liebe den Austausch mit euch! Austausch über und anhand von Literatur in Social Reading-Communities und auf Bücherblogs. In: Bartl, A., Behmer, M. (eds.) Die Rezension Aktuelle Tendenzen der Literaturkritik. pp. 225–252. Königshausen & Neumann, Würzburg (2017)
25. Ernst, T.: User Generated Content und der Leser-Autor als Prosumer. Potenziale und Probleme der Literaturkritik in Sozialen Medien. In: Kaulen, H., Gansel, C. (eds.) Literaturkritik heute Tendenzen - Traditionen - Vermittlung. V&R unipress, Göttingen (2011)
26. Trilcke, P.: Ideen zu einer Literatursoziologie des Internets. Mit einer Blogotop-Analyse. Textpraxis. 7(2), (2013)
27. Balijepally, V.G., Mangalaraj, G., Iyengar, K.: Are We Wielding this Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research. J. Assoc. Inf. Syst. 12(5), 375–413 (2011)
28. Elkan, C.: Using the Triangle Inequality to Accelerate k-Means. In: Proceedings of the 20th International Conference on Machine Learning, Washington D.C. (2003)
29. Punj, G., Steward, D.W.: Cluster analysis in marketing research: Review and suggestions for application. J. Mark. Res. 134–148 (1983)
30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)
31. Ward, J.H.J.: Hierarchical Grouping to Optimize an Objective Function. J. Am. Stat. Assoc. 58(301), 236–244 (1963)
32. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28(1), 101–108 (1979)
33. Zeng, D., Chen, H., Lusch, R., Li, S.: Social Media Analytics and Intelligence. IEEE Intell. Syst. 25(6), 13–16 (2010)
34. Guo, H., Jiang, Y.: Optimal Design of Consumer Review Systems. In: Proceedings of the 33d International Conference on Information Systems, Orlando (2012)
35. Babic Rosario, A., Sotgiu, F., de Valck, Bijmolt, T.H.A.: The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. J. Mark. Res. 53(3), 297–318 (2016)
36. Cheung, C.M.K., Thadani, D.R.: The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model. Decis. Support Syst. 54(1), 461–470 (2012)
37. Chevalier, J.A., Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. J. Mark. Res. 43(3), 345–354 (2006)

Less Complex than Expected – What Really Drives IT Consulting Value

Severin Oesterle[1], Arne Buchwald[2], and Nils Urbach[3]

[1] FIM Research Center, University of Bayreuth, Bayreuth, Germany
severin.oesterle@fim-rc.de
[2] Center for Digital Transformation (CDT), EBS Business School, Wiesbaden, Germany
arne.buchwald@ebs.edu
[3] FIM Research Center, University of Bayreuth, Project Group Business & Information Systems Engineering of Fraunhofer FIT, Bayreuth, Germany
nils.urbach@fim-rc.de

Abstract. Digitalization has a broad impact and the risk of external disruption is omnipresent throughout all industries which also applies to IT consulting firms. One response to this threat is to understand better the determinants of how value is created during the joint work on an IT project. Although previous literature offers valuable starting points for explaining value co-creation, no previous research synthesizes service provider and client perspective in a comprehensive model and empirically explains the co-creation of IT consulting service value. We build on the service-dominant (S-D) logic as the fundamental meta-theory and evaluate our deductively derived structural model based on 113 collected responses from IT consulting projects using structural equation modeling. Our major finding is that IT consulting service value only seems to be determined by consultant capabilities. Our findings provide new insights for S-D logic and service science literature and potential for future research.

Keywords: Value, Value Co-Creation, Service-Dominant Logic, Service Science, IT Consulting Services

1 Introduction

Digitalization has begun to change the rules of competition in many industries. The risk of disruption is omnipresent, and the digital transformation is on everyone's lips which is why organizations need to reconsider and to adjust their business models to remain competitive [1]. In this quest for digitalizing existing business models, creating new ones, or transforming existing services and processes, organizations often draw upon the knowledge of consultancies as external knowledge providers and transformation agents with specialized skills and resources to manage their digital transformation [2]. This way, organizations closely interact with IT consulting firms to achieve business value from such co-creations. In the same vein that regular organizations feel pressured to remain competitive, so do IT consulting firms [3]. While their fundamental business model has not changed for many decades, IT consulting firms are currently facing major

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

challenges. First, the IT consulting industry is strongly tied to the global economic situation, which means that in times of recession and resulting budget cuts at the clients' organizations, consulting assignments flatten out [2]. Second, IT consulting services are mostly labor intensive and require highly skilled and trained consultants. Particularly in the field of digitalization, IT consulting firms are in a 'war for talents'. Third, the billing model based on daily or hourly rates promotes overstaffed and protracted offerings which nowadays will be realized by clients through their increased experience with consulting offerings. Forth, the client organization's necessary shift to a more customer-centric view is also valid for consulting firms. The increasing experience of the clients with consulting services reinforces this major challenge once again [2]. Furthermore, the progress in digitalization paved the way for digital platforms (e.g., Comatch) in the consulting business that broker between supply (of highly qualified freelance consultants who formerly worked for large consultancies) and demand (organizations of different sizes requiring external advice). In addition, digital platforms offer organizations the opportunity to individually assemble their consulting team, so that the competition continues to increase [4].

To overcome the mentioned challenges, it is also for IT consultancies more essential than ever before to understand their customers' needs and IT literacy as well as to place the clients' requirement at the center of their activities and create value for them. Only a deeper understanding of which determinants drive the client's overall value enables ideal collaboration and co-creation settings and ultimately increases or at least maintains the IT consultancies' sales and profits and is beneficial to the client [2]. Therefore, both parties need to join their forces and individual capabilities. Surprisingly, empirical results on the underlying mechanisms between IT consulting firms' service provisions and client value are scarce. We connect to the research streams of service quality and customer satisfaction which is grounded in marketing literature and based on the distinction of technical and functional quality to measure customer satisfaction [e.g., 5, 6]. This research stream also includes approaches to measures service quality and satisfaction with the SERVQUAL instrument [7] and was adapted and used in information systems (IS) literature [e.g., 8, 9]. However, also the adapted IT-SERVQUAL instruments could not overcome the existing criticism [e.g., 10, 11].

The advent of the S-D logic can be considered as the most impacting shift in marketing and service science literature from a firm- and goods-centric to a customer-centric perspective [12-14]. Within the last years, research studies applied the S-D logic on various disciplines, e.g., service recovery [e.g., 15], service innovation [e.g., 16, 17-20], retailing [e.g., 21], tourism [e.g., 22], and IS [e.g., 23, 24-27]. Furthermore, S-D logic also impacted the service science research stream which is, in particular, relevant for us. The fundamental unit of analysis is service system, such as IT consulting service. Thus, we consider S-D logic as a meta-theory [25, 28] in which we theoretically ground our work, and we embed our work in service science.

We subsequently analyze existing quantitative-empirical studies in research streams which might be helpful for us. Although there are previous empirical IS studies applying SERVQUAL [e.g., 8, 9], they take an outdated firm- and goods-centricity and they focus on customer satisfaction (instead of value). Most literature on the meta-theory of S-D logic are conceptual contributions: First, the conceptual study of

Grönroos and Ojasalo [29] conceptually distinguish between customer-induced and provider-induced contributions to explain service productivity. While we second their suggestion that the service provider and the service recipient create the service in interaction, we disagree with their view that service productivity can be measured solely based on financial measures. Breidbach et al. [18] draw upon the S-D logic and investigate innovations in professional service firms, acknowledging a customer-centric perspective. Furthermore, Tallon [27] framed his study in a service science context and investigated the impact of business and IT strategy on firm performance. While both contributions provide valuable thoughts of value co-creation in a service system, their underlying objective differs.

The Work of Barrutia and Gilsanz [23] who examine electronic service quality and value in a business-to-consumer (B2C) e-commerce context inspired us. They state that "both consumer expertise and electronic service quality directly and positively affect value perception" [23, p.231] and incorporated a consumer and firm resources in their model. Whereas their model and the incorporation of consumer and form resources serve as a valuable starting point, the investigated B2C-service cannot be compared with the more complex IT consulting services. Concluding, all previous works neglect certain parts central to our research objective: they focus on B2C-relationships, the underlying objective differs, and/or the investigated services are not applicable and transferable to IT consulting services. Hence, no validated quantitative-empirical model synthesized service provider and client capabilities in a business-to-business (B2B) service system. To address this research gap, we investigate the following research question: Which factors determine the IT consulting service value considering both consultant and client capabilities?

2 Theoretical Foundations

We ground our research in the S-D logic of Vargo and Lusch [12] as "philosophical foundation" [28, p.19]. With the occurrence of S-D logic, the dichotomy of goods and services is overcome. Vargo and Lusch [12, p.2] define service "as the application of specialized competences […] through deeds, processes, and performances for the benefit of another entity or the entity itself". Hence, goods represent the distributing mechanism for services because service is the fundamental unit of exchange [12, 13]. Taking this into account, S-D logic assumes that "all economies are service economies" [12, p.10] and "enterprises cannot deliver value, but only offer value propositions" [13, p.7] which leads to an enhanced client role.

Moreover, S-D logic, as a meta-theory, can be seen as a theoretical foundation for service science [30]. Maglio and Spohrer [28, p.18] state that the "service-dominant logic may be the philosophical foundation of service science, and the service system may be its basic theoretical construct". Service science also focusses both on mutual interaction between the service provider and its customers in which specialized competences, such as knowledge and skills (operant resources), are exchanged and posits that a service "is a process of applying the provider's competences (knowledge and skills) for the benefit [value] of, and in conjunction with [value co-creation], the customer" [31, p.361]. In this framing, service systems are the basic unit of analysis and are defined "as value-co-creation configurations of people, technology, value propositions connecting internal and external service systems, and shared information (such as language, laws, measures, and methods)" [32, p.72].

2.1 Value Propositions, Value Co-Creation, and Value

The term value proposition is widespread in research as well as in practice. However, scholars realize that the term is yet poorly defined [17], also in the initial work and later revised works of S-D logic [14]. Lusch et al. [33] describe value propositions as a service provider's promise in which value-in-exchange is linked to value-in-use. Similarly, Grönroos and Voima [34, p.145] consider value propositions as "promises that customers can extract some value from an offering". However, the actual assessment of whether the service provision contributes to the client's value in future use can only be made by the client. Therefore, a service provider cannot guarantee an initial value contribution [17], but only make suggestions as to how the client can generate value through the use of the service. Barrutia and Gilsanz [23, p.232] which state that service provider and service recipient "simultaneously access, adapt, and integrate resources to create value for themselves and others" and assume that value is always co-created.

2.2 Consulting and Client Capabilities

Consulting service is defined as a time-limited achievement with the goal of defining, structuring, analyzing, and solving specific business problems interactively with the client [35]. Due to the interactive service provision and the knowledge-intensive industry and in line with Grönroos and Ojasalo [29], it is essential to consider both consulting and client capabilities in the value co-creation process. Furthermore, service science differentiates between operant (intangible, continuous, dynamic) and operand resources (tangible, static resources). We defined operant resources as those that act on operand resources [13]. Within consulting services, the fundamental unit of exchange is operant capabilities, such as knowledge and skills [36]. Thus, we solely focus on operant resources. To ensure a high consulting value, both parties should provide particular skills within the project. Both parties need certain types of expertise, such as social expertise, technological expertise, and functional expertise to maximize the consulting service value. In addition, consultants also need industry expertise and innovativeness to achieve high-quality results. However, IT consultancies can only provide value propositions which consist of their overall skill set. Thus, we integrate IT consulting quality as an operant resource in the consultant capabilities. IT consulting quality represents the overall perception of various underlying factors. Summarizing, the above outlined different types of expertise, as well as the consultant's innovativeness, contribute to the IT consulting quality which in turn contributes to the overall IT consulting service value.

Similar to the consultant capabilities, the client needs to provide a particular social, technological, and functional expertise. However, the client also needs specialized skills to handle the provided service [37]. A client who works on a time-limited project with consultants needs to understand how to work together with a consultant [38]. Hence, we consider the past experiences made while working with consultants as an operant client resource. Furthermore, an IT consulting service will only contribute to a client's value if the client has the ability to recognize the provided value and external information, as well as transform, assimilate, and apply it [39]. Thus, the client's required skills contribute to its absorptive capacity which in turn contributes to the overall IT consulting service value. Additionally, we hypothesize that the collaboration of both parties is an additional important determinant that contributes to the overall consulting service value and displays the co-creation of the consultant and the client. A prerequisite for consulting services is the exchange of knowledge and information between the consultant and the client as well as a trustworthy and courteous way of collaborating [40]. These social resources comprise for instance interpersonal trust, know-how exchange, relationship proneness, and social skills. We suggest that each of the parties' social expertise determines the collaboration quality which in turn directly influences the overall consulting service value.

3 Conceptual Development

After having described the theoretical foundations and identified previous works related to consulting service value, we now derive our hypotheses to explain the value co-creation between IT consultancies and clients. Our research aims at explaining and measuring consulting service value by both client and consultant capabilities. Therefore, we built on Oesterle et al. [41] in which we deductively derived a preliminary conceptual model. Whereas the previously published article served as a valuable starting point, we modified the previous model by including the concept of absorptive capacity [39] and by extending the model with additional antecedents. Investigating the value co-creation in the consulting industry, we focus on the project-level and examine the consulting service value that emerges from the mutual work of an IT consultancy and its client on a project level. As introduced in our theoretical foundation, we follow the distinction of Barrutia and Gilsanz [23] and integrate both client and consultant capabilities to capture the co-creation process within the consulting industry. We incorporate collaboration quality as an additional determinant of our dependent variable. Our dependent variable is the consulting service value which is the overall assessment of multiple factors (monetary and non-monetary) and emerges during the use of the provided service [23].

The first determinant that influences the consulting service value is collaboration quality which in turn is influenced by consultant and client capabilities. It refers to the extent to which at least two entities of the IT consultancy and the client worked jointly and coordinated together [42]. Thus, collaboration consists of personal interactions and relations between consultants and clients as well as collaborating aspects like respect and friendliness [43]. The better these qualities, the stronger are the ties between a consultant and its client, and thus, a higher value emerges [44]. Hence, we hypothesize:
H1: Collaboration quality has a positive impact on IT consulting service value.

3.1 Consultant Capabilities

The consultant should provide high consulting skills to address all relevant tasks in a structured and comprehensible way as well as practical research techniques applicable to the specific project. We define IT consulting quality as the extent to which a consultant has expert knowledge in required project skills such as systematic approach, statistical analysis, project and change management, development of surveys, or software engineering [45]. The necessary consulting skills can vary in each project. It is the consultant's task to assess which skill set is best for the project to achieve a high consulting service value. Hence, we state:

H2: IT Consulting quality has a positive impact on IT consulting service value.

Furthermore, the IT consulting quality depends on the consultants' industry knowledge. Consultants with high industry expertise better understand the specific needs of the client and have a thorough understanding of how business is conducted in the client industry [46]. We define industry expertise as the extent to which a consultant possesses expert knowledge in the domain of the client. We hypothesize:

H3: Industry expertise has a positive impact on IT consulting quality.

Moreover, consultants also need technological expertise. Especially in light of the ongoing digitalization, there are only a few consulting projects which do not include technology issues which stresses the importance of consultants possessing technological skills. The contracting of a consultant seems to be an easy way for clients to get access to new technologies. We define technological expertise as the extent to which a consultant possesses expert knowledge in technology and related areas [47] which facilitates the overall IT consulting quality. We hypothesize:

H4: Technological expertise has a positive impact on IT consulting quality.

Consultants also need functional expertise to complete consulting projects successfully. Consulting projects require a heterogeneous set of expert knowledge. We follow the definition of Hoffman [48, p.85] who defines a functional expert as "one who has special skills or knowledge derived from extensive experience with subdomains". Hence, we hypothesize:

H5: Functional expertise of the consultant has a positive impact on IT consulting quality.

Within the consulting industry, clients rely on consultants to figure out new ways of dealing with a specific issue. Especially in the light of the digitization of services, a certain level of innovative approaches is necessary. We define innovativeness as the degree to which consultants provide an innovative and novel service and are able to influence the client organization positively [49]. Hence, we hypothesize:

H6: Innovativeness of the consultant has a positive impact on IT consulting quality.

Finally, we define social expertise of the consultant as "interpersonal perceptiveness and the capacity to adjust one's behavior to different situational demands and to effectively influence and control the responses of others" [50, p.1076]. This conclusion also remains valid in a B2B context. Within a consulting project, there are various types of actors with different kinds of expertise and attitudes. Hence, the consultant must cope and collaborate with all of them to successfully complete the project. Thus, social expertise will facilitate the quality of the collaboration and furthermore contribute to

the IT consulting quality. Therefore, we hypothesize:
H7: Social expertise of the consultant has a positive impact on IT consulting quality.
H8: Social expertise of the consultant has a positive impact on collaboration quality.

Summarizing, we hypothesize that the determinants introduced above positively influence the client's perception of the overall IT consulting quality.

3.2 Client Capabilities

We now introduce the client capabilities through which the consulting service value emerges. Some determinants of the client capabilities are similar to the consultant capabilities but are assessed from the client's perspective (social, technological, and functional expertise). Also, we include the determinants experience with consultants and the client's absorptive capacity.

Absorptive capacity is a firm's ability to identify, assimilate, transform, and apply valuable external knowledge [39] which is also applicable to consulting services. The client has to be able to identify, assimilate, transform, and apply the consultant's external knowledge (as seen by the client) to create value. Therefore, we hypothesize:
H9: Absorptive capacity has a positive impact on IT consulting service value.

The client's social expertise, while similar to the consultant capabilities, takes the client's perspective. According to Ferris et al. [50, p.1076], social expertise is defined as the "interpersonal perceptiveness and the capacity to adjust one's behavior to different situational demands and to effectively influence and control the responses of others". In the same vein, client employees have to work jointly together with consultants. These skills will also help clients to conduct business with consultants jointly and absorb the provided solutions. Hence, we hypothesize:
H10: Social expertise of the client has a positive impact on collaboration quality.
H11: Social expertise of the client has a positive impact on absorptive capacity.

Like the consultant's technological expertise, also clients need to provide technological expertise. Otherwise, the client does not have the abilities to absorb the externally provided knowledge [47]. The client needs technological expertise to evaluate if the provided service is applicable and valuable. Therefore, we propose:
H12: Technological expertise has a positive impact on absorptive capacity.

Like the consultant's functional expertise, the client needs to understand all functional aspects of the project and provide expert knowledge. Due to the interactive service provision, the client requires these skills to assess if the externally provided service fits into the client company and is beneficial [51]. Hence, we conclude:
H13: Functional expertise has a positive impact on absorptive capacity.

Besides, we also integrate the determinant experience with consultants which we define as the extent to which client project members have developed empirical knowledge based on past interactions with consultants. This determinant is vital for clients since it reflects the learning process of how to interact, govern, judge, and transform the relationship with consultants [38]. Hence, we hypothesize:
H14: Experience with consultants has a positive impact on absorptive capacity.

4 Research Method

To analyze our model, we opted for quantitative-empirical methods for our research project because of their advantage regarding statistical generalization [52] and used structural equation modeling. The online questionnaire had the advantage that respondents were more likely to state their true unbiased opinion compared to a face-to-face interview. To empirically validate our theoretical model, we had to develop a suitable survey instrument.

4.1 Construct Operationalization

While existing measurement scales served as a good starting point, we had to adapt them in wording, language, and formality. In the cases that no suitable measurement scales were found or did not fit to our context, we developed additional items. After a first analysis of the initial item pool, we shortened our long-list and continued with the resulting item pool. After having adjusted the raw items in several rounds, we conducted two rounds of card-sorting procedures proposed by Moore and Benbasat [53] to assess the construct validity. In the first round, we achieved a satisfying overall hit ratio of 82.96%. However, the spread of the continuum was quite broad, so that we had to sharpen and to elaborate on some construct definitions and items. After a thorough revision of the items with the lowest scores and the highest cross-loadings and the construct definitions, we conducted a second round of card-sorting procedure. With an overall hit ration of 95.14% in the second card-sorting round, the score of the revised measurement model was satisfying.[1] After having successfully developed the measurement instrument, we conducted a pre-test with 20 participants (researchers, consultants, clients) which showed no need for further action. We measured all items with a seven-point Likert scale ranging from 'strongly disagree' to 'strongly agree'.

4.2 Data Collection

Having developed our survey instrument, we proceeded with the data collection. Our target population were consultants and employees who were jointly working on an IT project or an IT-related project. We used the convenience sampling approach when we approached potential participants in our professional business networks with an individual email including a personalized link for participation. The invitation contained a short description of our research endeavor as well as the assurance that we will handle all gathered data confidentially. The survey period lasted four weeks. The survey was provided in English to apply to a high number of participants. All participants answered the complete questionnaire and therefore, assessed own capabilities, the project partner's capabilities, the collaboration quality, and the consulting service value according to their perspective.

[1] We provide the final measurement model online: https://www.dropbox.com/s/e687qo8gcte6ja5/WI2019sub.pdf?dl=0

487 participants were invited, and 113 fully completed the questionnaire. Of the 113 participants, 67 were consultants and 46 clients and were mainly located in the German-speaking area. Most of the projects had one to five full-time equivalents and also one to five consultants working on the project. The client participants can be mostly assigned to higher management, and the consultant participants were mainly senior consultants or above. The client role within the projects can be considered as project managers and can be assigned to the area of IT. Most of the client participants were from the consumer goods industry (21.7%), followed by the financial service industry (19.6%), and manufacturing (15.2%). Regarding the consultancies, 59.7% were IT consultancies, followed by IT management consultancies (14.9%). 47.8% of the participants had a relationship above two years with their project partner.

5 Data Analysis and Results

First, to assess our measurement model, we tested for unidimensionality, internal consistency, indicator reliability, convergent validity, and discriminant validity [54, 55]. To measure unidimensionality, we conducted an exploratory factor analysis (EFA) using SPSS 20 with a principal component analysis in combination with the Varimax rotation. All of the identified factors had an Eigenvalue above 1.0 and loaded on their corresponding factor. We dropped one item of each of the constructs *functional expertise* and *absorptive capacity*, not meeting the threshold of 0.600 [56]. After these corrections, all factor loadings were above 0.600 and unidimensionality was shown. Furthermore, we tested our measurement model for internal consistency. We calculated Cronbach's Alpha (CA) and composite reliability (CR). Both CA and CR values of all constructs were above the threshold of 0.800, suggesting a high degree of internal consistency [57]. Furthermore, we checked our measurement model for indicator reliability. Therefore, all indicators should be greater than 0.707, and significant at the 0.05 level, which was the case for our data [57]. To assess convergent validity, we checked the average variance extracted (AVE), which should be higher than 0.500, which also applied to our data. Finally, we tested for discriminant validity. We calculated the Fornell-Larcker criterion [58], the heterotrait-monotrait ratio (HTMT) [59], and checked the cross-loadings [57]. Assessing the Fornell-Larcker criterion, all constructs had the highest correlation with themselves.[2] Furthermore, all constructs had an HTMT ratio below the threshold of 0.850 [59]. The cross-loadings also met the requirement to have the highest correlation with its corresponding construct. Hence, we conclude that the discriminant validity of our measurement model is given.

After having discussed the measurement model, we focused on the structural model by assessing path coefficients and coefficients of determination (R^2 values). To check our model for significance, we applied the Bootstrap procedure [57] with 5,000 subsamples. As we only considered completed questionnaires in our analysis, we did not have to cope with missing values. Figure 1 presents our findings showing R^2 values, path coefficients, and the level of significance.

---

[2] We provide the Fornell-Larcker Criterion online: https://www.dropbox.com/s/e687qo8gcte6ja5/WI2019sub.pdf?dl=0

Eight of our fourteen hypotheses could be confirmed on the 1%- or the 5%-level of significance and have weightings higher than 0.200 [57]. The coefficients of determination (R^2) indicated how much variance of a latent variable was explained by its preceding constructs [54]. Chin [57] distinguished between three groups of coefficients of determination (R^2): $R^2 \geq 0.670$ could be considered as substantial, $R^2 \geq 0.330$ as moderate, and $R^2 \geq 0.190$ as weak. Based on this classification, our structural model explained a moderate amount of variance of our dependent variable.

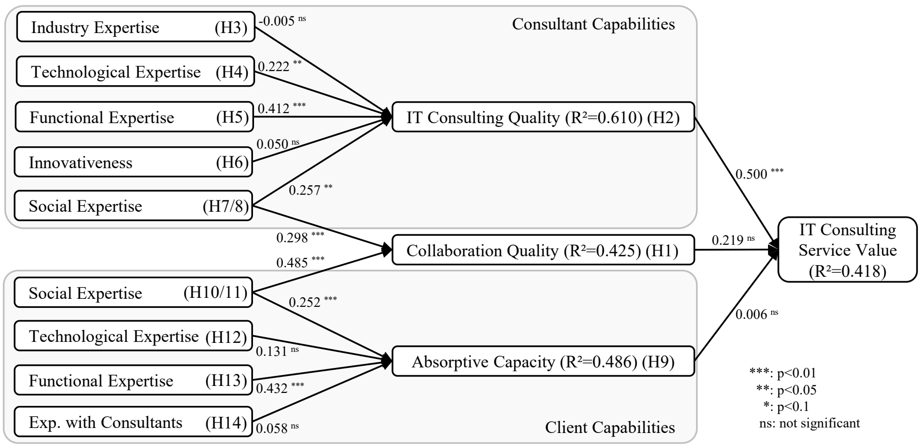


Figure 1. Assessment of the structural model

6 Discussion

Within the meta-theory of S-D logic and the service science research stream, there are numerous theoretical contributions and some qualitative-empirical studies. The number of quantitative-empirical studies which draw on the co-created emergence of value is surprisingly scarce. In this light, our major finding is that consulting service value only seems to be determined by consulting capabilities, while client capabilities, in contrast to our conceptualization, do not show any significant influence on consulting service value, neither in the overall data set nor in the subsamples. These results contradict the limited number of quantitative-empirical studies which suggest that service providers and their clients jointly co-create value. These studies, however, do not focus on a B2B service system [23], do not explain consulting service value but different dependent variables, such as loyalty, or do not focus on the consulting industry but other industries, such as hospitality [22]. As all these previous quantitative-empirical studies provide support for the co-creation of value within a B2C context, we argue that these service beneficiaries are always individual persons who have a high intrinsic motivation to co-create value and thus may have a higher priority to extract value from the service system. At a business level, the obtained value does not directly affect a tangible individual [60] which is why we propose that the client's intrinsic motivation in a B2B service system is lower than in a B2C service system. To the best of our knowledge, there is no previous quantitative-empirical study in a B2B service system that found support for the applicability of value co-creation.

Whereas IT SERVQUAL [8] as a measuring instrument for service quality relies on the gap model and focuses on how the service is delivered, our measurement model is not based on the disconfirmation paradigm. Thus, we concur with the various criticisms [10, 11] concerning both theoretical and operational aspects of the instrument. A second study is a conceptual model to explain customer satisfaction without an empirical validation [6]. Our empirical results provide support for their assumption that the functional expertise of the consultant positively contributes to service quality, which we included as IT consulting quality. IT consulting quality, as already briefly mentioned above, is a strong and significant determinant of consulting service value in our study, which means that IT consulting quality, such as addressing tasks in a structured and understandable way as well as providing the right project skills, contributes to the overall consulting service value. This result is in line with previous research [23]. Focusing in more detail on what IT consulting quality determines, our data show support for some of our hypotheses. We found support for our hypotheses that technological expertise (H4) and functional expertise (H5) are needed to provide a high IT consulting quality. Also, consultants need functional expertise to fully understand all facets of the project and apply their specialized skills to the actual problem. Hence, they can proceed their consulting competences. Next, we hypothesized that the IT consulting quality is influenced by the social expertise of the consultants (H7). The more a consultant possesses personal skills which help to direct its actions from an individual to a common action orientation, the higher is the IT consulting quality. Our data shows support for the hypothesis.

In contrast to our conceptualization, we also need to reject a few hypotheses. We hypothesized that consultants need to possess specific industry expertise that they leverage in projects for the success of the project, i.e., have a thorough understanding of how business is conducted in the client industry (H3). Our data shows no support for this hypothesis, and we reason that consultants and clients consider themselves smart enough to work in any given industry and context so that industry expertise is not viewed as a relevant determinant of IT consulting quality. Furthermore, we hypothesized that consultants need a certain degree of innovativeness to be able to provide unique and problem-solving solutions to the client (H6), which, however, is not supported by our data. This result is surprising, and we can only assume that there was no need for highly innovative solutions in the projects in our dataset or the client wanted standardized solutions.

Finally, we hypothesized that collaboration quality positively influences consulting service value (H1), which we cannot confirm. This finding is quite surprising. We posit that the value of the consulting service is determined by the mutual interaction and hence, the co-creation within the service provision. Especially in knowledge-intensive industries such as the consulting industry, mutual interactions of the service provider and the service recipient concerning information exchange and communication openness seem to be important. However, contrary to our hypothesis and based on the empirical findings, the relationship between collaboration quality and consulting service value is not supported. We assume that this non-significant relationship (H1) depends on the incorporated different types of consulting project types. Whereas consultants can provide some IT consulting projects with a low collaboration level (e.g., IT strategy projects, IT audit), some others need a high collaboration level (e.g., IT implementation, IT transformation).

Summarizing, our data set does not support one of the main assumptions of S-D logic, the co-created determination of value in the domain of IT consulting services. It seems as if the consulting industry is still service provider-centric, not only in the perception of the consultants but also in the perception of the client. Our findings show consulting capabilities solely determine the consulting service value.

7 Conclusion

With our work, we set out to empirically identify determinants which explain the emergence of co-created value in consulting relationships from both the perspective of IT consultancies and clients. To achieve this objective, we adopted and further developed a conceptual model [41]. Our key contributions are twofold. First and foremost, our study showed that consulting service value is determined by consultants only, neglecting any influence of the client. Second, the consultant's functional, technological, and social expertise mainly predict IT consulting quality.

Before we conclude this paper by outlining our recommendations for future research and by highlighting our contributions to both theory and practice, we briefly discuss our study's limitations. Our study only focuses on perceived value, which can be considered as a key determinant of consulting service success. Success in that respect may also be influenced by additional factors such as price, political connections, and sales capabilities which, however, is beyond the study's scope. Furthermore, the participants completed the whole survey, i.e., also the assessment of own capabilities which could be biased, and thus, participants might have overestimated them. This bias could be minimized by collecting matched pairs.

Regarding the next steps, we will follow a dyadic data analysis approach [61] to overcome the possibly biased self-assessment of the participants. Within the dyadic analysis, clients will evaluate their consultants and vice versa on a project level. Through this measurement, it will be possible to get an objective judgment of the incorporated consulting and client capabilities. Thus, we will obtain more profound insights. To account for the particularities of the consulting domain, we aim to strengthen our statistical analysis by carrying out multi-group comparisons. Especially, multi-group comparisons regarding the different IT consulting project types may lead to new insights on the underlying mechanisms.

Keeping the limitations in mind, our results contribute to both theory and practice. Our contribution to the research stream of service science is the advancement of the theoretical discourse on value emergence by providing an empirically validated theory. Furthermore, we account for the value co-creation model in a service system which is the fundamental basis of analysis in service science. Our study indicates that value co-creation in an IT consulting service system is still service provider centric. The fact that our data do not support the co-creation of value provides new insights into a service system as well as foster the discussion of the meta-theory S-D logic.

From a practical point of view, we provide meaningful insights to both consultant and client organization. IT consultancies may learn which determinants lead to a high consulting service value and can be used internally for human resources development purposes. In addition to consultancies, our results equip clients with information which determinants matter most in the selection of a consultancy to maximize the value they gain from contracting a consultancy.

References

1. Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J.M., Loos, P., Spann, M.: Business Models. Bus Inf Syst Eng 6, 45-53 (2014)
2. Poulfelt, F., Olson, T.H.: The Changing Global Consulting Industry. In: Poulfelt, F., Olson, T.H. (eds.) Management Consulting, vol. 2, pp. 5-36. Routledge, New York (2018)
3. Nissen, V., Seifert, H.: Digital Transformation in Business Consulting. In: Nissen, V. (ed.) Digital Transformation of the Consulting Industry, pp. 153-190. Springer, Cham (2018)
4. Nissen, V.: Consulting Research: A Scientific Perspective on Consulting. In: Nissen, V. (ed.) Advances in Consulting Research, pp. 1-27. Springer, Cham (2019)
5. Grönroos, C.: A Service Quality Model and its Marketing Implications. European Journal of Marketing 18, 36-44 (1984)
6. Das, A., Soh, C.W.L., Lee, P.C.B.: A Model of Customer Satisfaction with Information Technology Service Providers. Proceedings of the Conference on Computer Personnel Research 190-193 (1999)
7. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A Conceptual Model of Service Quality and Its Implications for Future Research. Journal of Marketing 49, 41-50 (1985)
8. Yoon, S., Suh, H.: Ensuring IT Consulting SERVQUAL and User Satisfaction: A Modified Measurement Tool. Information Systems Frontiers 6, 341-351 (2004)
9. Kettinger, W.J., Lee, C.C.: Pragmatic Perspectives on the Measurement of Information Systems Service Quality. MIS Quarterly 21, 223-240 (1997)
10. Buttle, F.: SERVQUAL: review, critique, research agenda. European Journal of Marketing 30, 8-32 (1996)
11. Dyke, T.P.v., Leon, A.K., Prybutok, V.R.: Measuring Information Systems Service Quality. MIS Quarterly 21, 195-208 (1997)
12. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. Journal of Marketing 68, 1-17 (2004)
13. Vargo, S.L., Lusch, R.F.: Service-Dominant Logic: Continuing the Evolution. Journal of the Academy of Marketing Science 36, 1-10 (2008)
14. Vargo, S.L., Lusch, R.F.: Institutions and Axioms: An Extension and Update of Service-Dominant Logic. Journal of the Academy of Marketing Science 1-19 (2016)
15. Dong, B., Evans, K., Zou, S.: The effects of customer participation in co-created service recovery. Journal of the Academy of Marketing Science 36, 123-137 (2008)
16. Lusch, R.F., Nambisan, S.: Service Innovation: A Service-Dominant Logic Perspective MIS Quarterly 39, 155-175 (2015)
17. Skålén, P., Gummerus, J., von Koskull, C., Magnusson, P.R.: Exploring Value Propositions and Service Innovation. Journal of the Academy of Marketing Science 43, 137-158 (2015)

18. Breidbach, C.F., Smith, P., Callagher, L.J.: Advancing Innovation in Professional Service Firms: Insights from the Service-Dominant Logic. Service Science 5, 263-275 (2013)
19. Barrett, M., Davidson, E., Prabhu, J., Vargo, S.L.: Service innovation in the digital age: key contributions and future directions. MIS quarterly 39, 135-154 (2015)
20. Breidbach, C.F., Maglio, P.P.: A Service Science Perspective on the Role of ICT in Service Innovation. Proceedings of the 23rd European Conference on Information Systems, paper 33 (2015)
21. Karpen, I.O., Bove, L.L., Lukas, B.A., Zyphur, M.J.: Service-Dominant Orientation: Measurement and Impact on Performance Outcomes. J of Retailing 91, 89-108 (2015)
22. Blazquez-Resino, J.J., Molina, A., Esteban-Talaya, A.: Service-Dominant Logic in tourism: the way to loyalty. Current Issues in Tourism 18, 706-724 (2015)
23. Barrutia, J.M., Gilsanz, A.: Electronic Service Quality and Value. Journal of Service Research 16, 231-246 (2013)
24. Alter, S.: Does Service-Dominant Logic Provide Insight about Operational IT Service Systems? In: AMCIS 2010 Proceedings, paper no. 138. AIS, (2010)
25. Brust, L., Breidbach, C., Antons, D., Salge, T.-O.: Service-Dominant Logic and Information Systems Research. Proceedings of the 38th International Conference on Information Systems (2017)
26. Kutsikos, K., Sakas, D.: A Framework for Enabling Service Configuration Decisions. Procedia - Social and Behavioral Sciences 148, 604-610 (2014)
27. Tallon, P.P.: A Service Science Perspective on Strategic Choice, IT, and Performance in U.S. Banking. Journal of Management Information Systems 26, 219-252 (2010)
28. Maglio, P.P., Spohrer, J.: Fundamentals of service science. Journal of the Academy of Marketing Science 36, 18-20 (2008)
29. Grönroos, C., Ojasalo, K.: Service productivity: Towards a conceptualization of the transformation of inputs into economic results in services. Journal of Business Research 57, 414-423 (2004)
30. Vargo, S.L., Lusch, R.F., Akaka, M.A.: Advancing Service Science with Service-Dominant Logic. In: Maglio, P.P., Kieliszewski, C.A., Spohrer, J.C. (eds.) Handbook of Service Science, pp. 133-156. Springer, New York (2010)
31. Chew, E.K.: Service Science. In: Maglio, P.P., Kieliszewski, C.A., Spohrer, J.C. (eds.) Handbook of Service Science, pp. 359-386. Springer, New York (2010)
32. Spohrer, J.C., Maglio, P.P., Bailey, J., Gruhl, D.: Steps toward a science of service systems. Computer 40, 71-77 (2007)
33. Lusch, R.F., Vargo, S.L., O'Brien, M.: Competing through Service: Insights from Service-Dominant Logic. J of Retailing 83, 5-18 (2007)
34. Grönroos, C., Voima, P.: Critical Service Logic: Making Sense of Value Creation and Co-creation. Journal of the Academy of Marketing Science 41, 133-150 (2013)
35. Nissen, V.: Consulting Research. In: Nissen, V. (ed.) Consulting Research, pp. 3-38. DUV, Wiesbaden (2007)
36. Chandler, J.D., Vargo, S.L.: Contextualization and Value-in-Context: How Context Frames Exchange. Marketing Theory 11, 35-49 (2011)
37. Løwendahl, B.R., Revang, Ø., Fosstenløkken, S.M.: Knowledge and Value Creation in Professional Service Firms. Human Relations 54, 911-931 (2001)
38. Gentile, C., Spiller, N., Noci, G.: How to Sustain the Customer Experience. European Management Journal 25, 395-410 (2007)
39. Cohen, W.M., Levinthal, D.A.: Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly 35, 128-152 (1990)

40. Wurst, K.M., Hoegl, M.M., Gemuenden, H.G.: Collaboration within and between teams in multi-team R&D projects. Proceedings of the Portland International Conference on Management of Engineering and Technology 545-552 (2001)
41. Oesterle, S., Buchwald, A., Urbach, N.: Understanding the Co-Creation of Value Emerging from the Collaboration between IT Consulting Firms and their Customers. Proceedings of the 37th International Conference on Information Systems (2016)
42. Pereira, C.S., Soares, A.L.: Improving the Quality of Collaboration Requirements for Information Management through Social Networks Analysis. International Journal of Information Management 27, 86-103 (2007)
43. Kelley, S.W., Donnelly Jr, J.H., Skinner, S.J.: Customer Participation in Service Production and Delivery. J of Retailing 66, 315 (1990)
44. Boughzala, I., de Vreede, G.-J.: Evaluating Team Collaboration Quality. Journal of Management Information Systems 32, 129-157 (2015)
45. Boh, W., Ren, Y., Kiesler, S.: Managing expertise in a distributed environment. Proceedings of the 23rd International Conference on Information Systems 173-181 (2002)
46. Goles, T.: Vendor capabilities and outsourcing success: A resource-based view. Wirtschaftsinf 45, 199-206 (2003)
47. Kirby, D.A., Dylan, J.-E.: Small Technology-based Professional Consultancy Services in the United Kingdom. The Service Industries Journal 17, 155-172 (1997)
48. Hoffman, R.R.: How can Expertise be Defined? Implications of Research from Cognitive Psychology. Exploring Expertise 81-100 (1998)
49. Garcia, R., Calantone, R.: A critical look at technological innovation typology and innovativeness terminology. J of Product Innovation Management 19, 110-132 (2002)
50. Ferris, G.R., Witt, L.A., Hochwarter, W.A.: Interaction of social skill and general mental ability on job performance and salary. J of Applied Psychology 86, 1075-1082 (2001)
51. Richter, A., Niewiem, S.: Knowledge transfer across permeable boundaries. Scandinavian Journal of Management 25, 275-288 (2009)
52. Johnson, P., Duberley, J.: Understanding management research: An introduction to epistemology. Sage, London (2000)
53. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. ISR 2, 192-222 (1991)
54. Hair, J.F., Hult, G.T.M., Ringle, C., Sarstedt, M.: A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publications (2016)
55. Urbach, N., Ahlemann, F.: Structural equation modeling in information systems research using partial least squares. JITTA 11, 5-40 (2010)
56. Gefen, D., Straub, D.: A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. CAIS 16, 5 (2005)
57. Chin, W.W.: The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (ed.) Modern methods for business research, pp. 295-336. Lawrence Erlbaum Associates, Mahwah, New Jersey, USA (1998)
58. Fornell, C., Larcker, D.F.: Structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18, 382-388 (1981)
59. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science 43, 115-135 (2015)
60. Roser, T., DeFillippi, R., Samson, A.: Managing your co-creation mix: co-creation ventures in distinctive contexts. European Business Review 25, 20-41 (2013)
61. Kenny, D.A., Kashy, D.A., Cook, W.L.: Dyadic Data Analysis. The Guilford Press, New York, London (2006)

Modularity Canvas – A Framework for Visualizing Potentials of Service Modularity

Jens Poeppelbuss[1] and Alexander Lubarski[2]

[1] Ruhr-Universität Bochum, Lehrstuhl für Industrial Sales and Service Engineering, Bochum, Germany
jens.poeppelbuss@rub.de
[2] University of Bremen, Business Studies & Economics, Bremen, Germany
lubarski@uni-bremen.de

Abstract. Service modularity has been proposed as a possible solution to the dilemma between the customer-driven thirst for individualization and the standardization ambitions of service providers. However, current modularization methods that intend to support the corresponding transformation process remain mostly on a conceptual level with little real-life application. Therefore, we develop the Modularity Canvas – a generic framework for the structured information capturing and identification of potentials for advancing towards a modular service architecture. The artifact is developed following the Design Science Research Methodology. The results include the artifact itself and first insights from five demonstration workshops at industrial service providers from contract logistics, wind energy and automotive engineering sectors. The contribution of the paper lies in the deeper understanding of what dimensions need to be considered when preparing service providers for their move towards a modular service architecture.

Keywords: Service modularity, canvas, framework, modularization method, design science

1 Introduction

In business-to-business (B2B) markets, customer requirements are highly individualized and complex, meaning that the same service value proposition cannot be offered twice to different customers without profound alteration [1]. However, growing competitive pressure also forces B2B service providers to constantly search for cost efficiency improvements by standardizing their resources and processes. In this regard, the concept of service modularity has been proposed as a viable solution to the dilemma between individualization and standardization [2].

A modular system is comprised of smaller parts (i.e., modules) with standardized interfaces and clearly defined functionalities, features, or values, which can be designed, improved and exchanged independently, yet function together as a whole [3]. This way, the provider can try to maximize external variety (i.e., the number of possible variants due to configuration possibilities), while minimizing internal variety

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

(i.e., a finite and manageable amount of service elements or modules). Although the concept of modularity has been successfully established in the context of software engineering [4] and manufacturing [5], recent academic discussions on service modularity have often remained on a conceptual level with little evidence for practical adoption [6]. Although there exist several service modularization methods in the academic literature [7], it seems that these are not acknowledged by practitioners, which can be partly ascribed to their abstract and conceptual nature. Nevertheless, it is especially the field of services, in which the concept of modularity is expected to further accelerate [8, 9]. Moreover, recent works also conceptualize service innovation as a "recombination of resources" [10, p. 380] and strongly rely on ideas related to the concept of service modularity.

Against this background, the objective of our research is to develop a framework that service providers can use to create visual models of their status quo in service provision and to identify potentials for developing towards a modular service architecture. We name this framework the Modularity Canvas. Similar to the very popular Business Model Canvas, the Modularity Canvas intends to offer a "shared language for describing, visualizing, assessing, and changing" [11] service modularity in organizations. Our development process follows the Design Science Research Methodology (DSRM) proposed by Peffers et al. [12]. The paper presents our results after the first four stages of the DSRM (namely *1. Problem Identification*, *2. Objective Definition*, *3. Design and Development*, and *4. Demonstration*), leaving the further stages of *5. Evaluation* and *6. Communication,* as well as possible design iterations, for our upcoming research activities. The Modularity Canvas focuses on the very first phase of information capturing at the beginning of a modularization initiative, before service modules are defined. This phase has been identified as being under-researched and not well covered with method support [7]. The theoretical contribution of the paper lies in the deeper understanding of the dimensions (both on a strategic and operational level) that need to be considered when preparing service providers for their move towards a modular service architecture.

The remainder of this paper is structured as follows. Next, we give a brief background on service modularity, canvases, and ontologies, followed by a section on our research approach. We present the artifact in Section 4 and report on our demonstration workshops in Section 5. Section 6 gives a summary and an outlook.

2 Research Background

2.1 Service Modularity

The research on service modularity is still considered to be in its infancy [6]. Pekkarinen and Ulkuniemi [2] were among the first to apply the concept of modularity to business services and highlighted the multi-dimensional nature of service modularity. Ever since, service modularity has attracted the attention of researchers from different academic fields such as marketing, information systems (IS), engineering, and even psychology [6]. The expected benefits of service modularity include, amongst others, efficiency benefits for the service provider [13], economies of scale and scope [14], the reusability of service elements for future offerings [15], as well as flexibility and faster development cycles [16]. Besides provider-related benefits, modular customization is also believed to directly improve the service experience and the loyalty of customers [17]. To leverage such benefits, Carlborg and Kindström [15] outline different modular strategies based on a distinction between four service types with different process characteristics (rigid vs. fluid) and different customer roles (passive vs. active).

In order to support the transformation process of a service provider from a rather monolithic towards a modular service architecture, scholars have been focusing on the design of appropriate methods. Depending on the underlying service paradigm [18], these methods were either transferred and adjusted from the domain of products [19], developed specifically for services [16], or reflect a simultaneous modularization of products and services due to their inseparability and joint value-creation [20]. A synopsis of existing methods for service modularization is provided by Poeppelbuss and Lubarski [7], who systematize them along, first, an idealistic modularization process (Figure 1) and, second, different types of modular structures. Further pertinent works include the iterative guideline FAMouS for architecting modular services [21], and a list of three trade-offs that are to be tackled when translating modularity into a functional set of design choices for professional services [22].

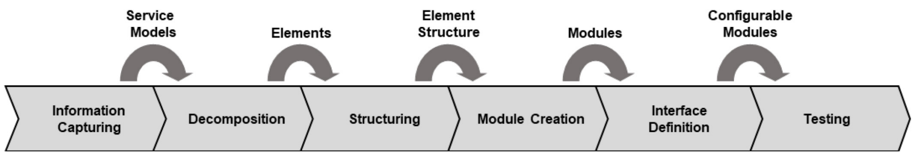


Figure 1. Phases of the modularization process [7]

However, there is little empirical evidence that these strategies and methods have actually been adopted in practice; apart from information given about the testing environments and case studies presented within the papers themselves [5]. In fact, very little is known about whether and how practitioners actually conceptualize and document modular service architectures at all. We see three explanations for this issue. First, the majority of existing modularization methods oftentimes make unrealistic assumptions or define ambitious prerequisites such as the existence of an already decomposed service portfolio or a clear assignment of required resources to operational service processes. Second, due to the academic origin of many modularization methods, they tend to remain on an abstract level and provide rather generic recommendations only. For instance, the FAMouS framework for architecting modular services [21] does not provide practical details on how to actually perform the steps such as “identification of elements” or “analysis of elements”. Third, existing works mostly approach service modularization from the strategic perspective of variety management and disregard the interrelationships with operational sales processes and the quotation process in particular. IT requirements for implementing and communicating a modular service portfolio in marketing and sales activities are hardly discussed, although modular architectures can enable the use of online configurators and so-called configure-price-quote (CPQ) software. Hence, we assume that practitioners cannot determine easily how their organization can benefit from the concept of service modularity and, more importantly, where to start with their own modularization initiative.

2.2 Canvases and Ontologies

Inspired by the worldwide success and great acceptance of the Business Model Canvas (BMC) [11], both in academia and practice and across various disciplines, we decided to use a canvas representation for visualizing the intended framework. A canvas representation is a concise, easy-to-understand, easy-to-use, and visually appealing overview of the key components required for a specific subject area. One of the first canvases developed in an academic setting was the Strategy Canvas evaluating a company's position within its respective industry based on pre-defined factors of competition [23]. Almost a decade later, the BMC became a role model of how a canvas can have a major impact on both the theoretical discussion and practical application of designing business models. Ever since its introduction, it has been used by various consulting agencies and was further extended and modified to fit specific purposes, including the Service Business Model Canvas [24], for instance.

The BMC was initially presented as a business model ontology [25]. Following Uschold and Gruninger [26], an ontology provides a shared understanding and conceptualization of a domain of interest that can be used as a unifying framework to facilitate knowledge sharing and re-use, as well as inter-operability between different organizational entities and systems. An ontology provides a conceptual framework for modeling domain knowledge and it is specified in the form of definitions of representational vocabulary. While Osterwalder [25] also worked on a formal representation of business models with an own XML-based language, the BMC, as it is known today, mainly offers an informal framework, which is to a large extent "expressed loosely in natural language" [26, p. 6].

Similarly, the Modularity Canvas that we present in this paper is supposed to help organizations in creating a shared understanding of their current service architecture and to derive actions to achieve a (more) modular service architecture thereof. Opposed to previous frameworks that cover the whole modularization process, it mainly addresses the initial phase of modularization initiatives in organizations. Hence, the Modularity Canvas covers the *Analysis* phase of the FAMouS framework of Dörbecker and Böhmann [21] and the *Information Capturing* phase of the classification framework of Poeppelbuss and Lubarski [7] (Figure 1), both of which emphasize the key role of the initial phases for the overall modularization initiative.

3 Methodology

Consistent with Osterwalder [24], we consider the development of the Modularity Canvas and its underlying ontology to be subject of design science and, hence, follow a corresponding research approach. Following the distinction between different artifact types (constructs, model, method, and instantiation) by March and Smith [27] and Osterwalder's [24] categorization, the different fields of the Modularity Canvas can be considered constructs, and the actual canvas is a model (i.e., the ontology). Our research process follows the DSRM by Peffers et al. consisting of six phases [11]:

1. Identify Problem and Motivate: We extracted the research problem and the motivation for the development of the Modularity Canvas from the current literature on service modularity (section 2) as well as through a pre-study with representatives (mainly managing directors and heads of sales) from industrial services providers in the contract logistics and wind energy sectors [28]. In this pre-study, we conducted 17 semi-structured expert interviews. The interview results show that most of the providers face challenges to display the full spectrum of their services in a structured and understandable way to their customers. This is partly due to the heterogeneity of their service offerings, but also due to the concerns of losing the personal touch towards their customers. However, the majority of the interviewed partners also sees the quotation process (i.e., the preparation of the quotation document containing how and at what price the provider will solve the customer's problem) as a current pain point and a possible area for improvement. They indicate that the preparation of the quote is often made from scratch costing too much effort and time. They further point to the issue of not well-integrated IT systems and a proliferation of spreadsheets, for which the management of master data is obviously a problem. From the interviews, we generally see a relatively poor adoption of the service modularity concept in practice. The experts confirm that the conceptual ideas and methods from academia are hardly known and not used. Therefore, we identify the need for a lightweight and easy-to-use tool in the form of a canvas, as the BMC has shown that a tool of this kind can be valuable to both practice and academia and achieve widespread adoption. As the Modularity Canvas must especially support the early phase of modularization initiatives, which is not well covered by existing method support [7] yet, we also consider a canvas representation as particularly useful. It can guide workshop discussions, which are common in the early phases of such initiatives.

2. Define Objectives of a Solution: As the Modularity Canvas is supposed to support the early phase of modularization initiatives; it is required to support the capturing and visualization of information about the status quo, the search for modularity potentials, and the definition of directions for improving variety management and the quotation process. It is not supposed to support the actual definition of service modules, for which many methods already exist [7]. As the canvas will be used in workshops with board markers and sticky notes, it must offer structure and flexibility alike. This means that it has to provide guidance during discussions and to direct the participants' focus to important aspects and their interrelationships, but also to offer the freedom of annotating notes wherever needed. Further guidance should also be provided for the interpretation of the canvas (e.g., how to identify potentials for designing modular service offerings from the status quo as depicted in the canvas).

3. Design and Development: The findings from our pre-study [28] indicate that there are two layers of service modularity that need to be considered, which include the strategic variety management and the operational quotation activities in sales.

Hence, the Modularity Canvas and its underlying ontology must correspond to these two layers. In order to further populate the two layers, we relied on existing academic literature (e.g., with regard to the dimensions of services offerings) and concepts prevalent in practice (e.g., configure, price, quote, CPQ). We give the detailed reasoning for the corresponding fields in the artifact description in the next section.

4. Demonstration: For demonstrating its usability, we utilized the Modularity Canvas in workshops with five German industrial service providers from contract logistics, wind energy and automotive engineering sectors. We gathered qualitative feedback from the participants concerning the workshops, the Modularity Canvas as a central tool of the workshops, and the documentaries that we provided to them after post-processing the workshop results. We were provided with valuable insights that we will address in a further design iteration of the Modularity Canvas.

This paper presents the current version of the artifact as used in the demonstration workshops. The DSRM further includes the two steps of *5. Evaluation* and *6. Communication*, which we plan to conduct in our upcoming research activities.

4 The Modularity Canvas

The Modularity Canvas provides a framework that organizations can use to create visual models for describing the rationale of how they manage service variety and modularity on a strategic level, and how they organize their corresponding operational quotation process. It comprises eleven fields that correspond to the underlying ontology of service modularity, which we will justify in the following.

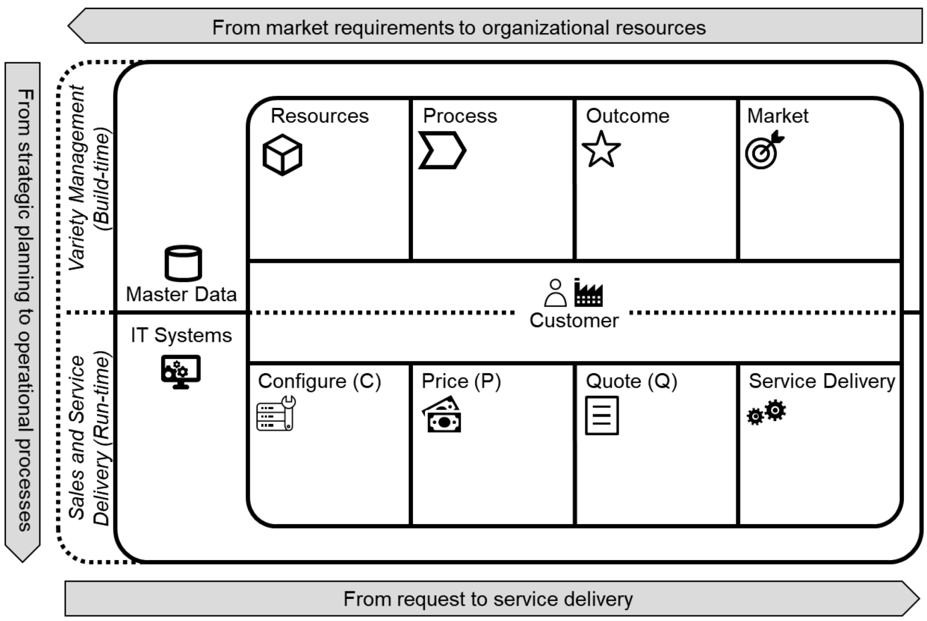


Figure 2. Modularity Canvas

The canvas representation of this ontology (i.e., the Modularity Canvas; Figure 2) comprises two layers: The upper layer focuses on *Variety Management*, which reflects strategic decisions about the modular service architecture made at *build-time*. The lower layer comprises the operational *Sales and Service Delivery* processes at *run-time*. The central field of *Customer* connects the two layers. They are surrounded by the two fields of *IT Systems* (supporting the operational processes) and *Master Data* (defining the service catalog or architecture and its elements).

The recommended order for capturing the status quo about the service architecture in workshops is starting with the strategic layer in the upper right corner, moving from market requirements to organizational resources, moving from the upper to the lower layer, and then moving from the request for quote to service delivery. However, if issues in the operational quotation process are the actual trigger for the modularization initiative, it can also be meaningful to start with the lower, operational layer first. Table 1 gives examples of guiding questions that are supposed to stimulate the workshop discussions when filling out the Modularity Canvas.

The *upper layer* is structured along the four steps towards a strategic service vision according to Heskett [29] and Pekkarinen and Ulkuniemi [2]. These four steps also correspond to four dimensions frequently used to describe service offerings [30, 31]:

1. Identification of target market segments (*Market* dimension),
2. Development of service concepts to address targeted markets' needs (*Outcome* dimension),
3. Codification of an operating strategy to support the service concept (*Process* dimension), and
4. Design of a service delivery system to support the operating strategy (*Resource* dimension).

The *Market* dimension depicts in how far the organization promotes its service offerings to specific market segments or even individual customers instead of an undifferentiated (one-size-fits-all) approach. Becker et al. [30] motivate the need for a distinct market dimension by the fact that service offerings require value co-creation with customers, and, thus, a particular fit with their specific demands. In our canvas, this field is supposed to capture in how far marketing communications and sales activities differentiate between market segments and even individual customers (e.g., through different customer interfaces and sales channels [2]). Correspondingly, the range of different interfaces and channels as well as the set of approaches used for market segmentation and customer insight analytics are noted down here.

The *Outcome* dimension captures the value proposition that the organization offers, which is typically represented by service concepts or product models [30, 31]. Such service concepts and product models comprise a definition of the service contents and a superordinate structure in terms of a service catalog or portfolio [31]. Here, the canvas captures the range of services that are offered and in how far the service portfolio has some structure, or even a modular architecture. The traditional literature on product modularity typically centers around this dimension when parts or components of a product are divided into modules that can be easily interchanged and replaced, while service modularity has been recognized as a more complex, i.e., multi-dimensional concept [2].

The *Process* dimension focuses on the activities performed by the service provider as well as the customer as an external factor to generate the outcome. “Whereas product models map *what* a service does, process models describe *how* the outcomes of a service are achieved” [31]. Here, the canvas captures in how far service delivery processes are broken down into standardized or customized sub-processes and how the process design allows for quick and flexible responses to varying customer requirements [12].

Table 1. Examples of guiding questions

Canvas field	*Guiding question*
Market	How are different customer segments addressed with service variants?
Outcome	Which service variants exist?
Process	Which process variants exist?
Resource	Which different resources are used?
Master Data	How are the variants represented in the master data?
Customer	How does the customer influence the variance of marketing activities, outcome, processes and resources? What interactions with the customer happen during the quotation process?
IT Systems	Which application systems support the quotation process?
Configure (C)	How is the service offering composed as a response to a request for a quote?
Price (P)	How is the price of a bundle of service modules determined?
Quote (Q)	How are quotation and contract documents designed?
Service Delivery	To what extent is information exchanged between sales and service delivery?

The *Resource*[1] dimension depicts the configuration of the service delivery system. The service delivery system defines how the resource base is organized, including human resources, equipment, organizational units, and supply chains for service delivery; and in how far the resource configurations can be flexibly adjusted depending on specific service cases. In this regard, Pekkarinen and Ulkuniemi [2] refer to the term “organizational modules” that provide standardized ways to organize a service provider’s internal and external resources for maximum efficiency. As examples for internal resource modules, they identify teams for specific customer segments or competence areas. Examples for external organizational modules include subcontracting, the use of hired labor, and alliances.

[1] With regard to this dimension, Bullinger et al. [31] refer to a “structure dimension”, whereas Becker et al. [30] name it “potential dimension”. In their textual explanations, however, both refer to the *resources* required to deliver a service or value bundle, so we opted for the term *Resource* dimension.

The decisions made on this upper layer are typically reflected by the *Master Data* in the organization's IT systems, including customer master data (Market), service and product master data, catalogs, or portfolios (Outcome), process definitions and job instructions (Process), as well as employees, assets and further organizational master data (Resource). The corresponding field in the canvas is supposed to capture what master data is available and how it is managed (including the collection, quality assurance, and distribution throughout the organization).

The master data is used by *IT Systems* to support the operational activities at the lower layer, which typically involve online shops and configurators, customer relationship management (CRM) and enterprise resource planning (ERP) systems, e-mail applications, as well as spreadsheets and desktop tools. The systems in use are captured in the corresponding field of the canvas. As the two fields *Master Data* and *IT Systems* surround the other fields on the two levels, it is recommended to depict which master data is available for which service dimension and which system is used for which phase of the lower layer.

The phases on the *lower layer* are also structured into four fields. The first three fields on the *lower layer* comprise the configuration (C), pricing (P) and quote generation (Q) activities that accompany the operational sales process, which can be supported by so-called CPQ software applications [32].

In the *Configure* phase, the service offering is specified or configured based on customer needs. This can happen from scratch, by adjusting similar specifications from the past, or by already using available service modules. The activities in this phase are typically carried out by a sales clerk relying on her/his discussions with customers about their needs. The feasibility of the specification/configuration has to be assessed and ensured, which may also lead to discussions with internal experts. In case of online configurators, the user can define her/his configuration based on existing modules through a web-based frontend as a self-service.

In the *Price* phase, the sales clerk calculates a price for the specification/ configuration of the service and additional parameters using the pricing engine of a software (or simply a spreadsheet tool, which is very common in practice). Depending on the industry and market environment, the pricing strategies to be implemented by the pricing engine can vary from rather simple cost-based and linear strategies, where the price of the specification/configuration can be determined from the prices as defined for single service components or modules, to more complex competition-oriented and demand-oriented pricing strategies.

In the *Quote* phase, a document with the configuration and the calculated price is generated, which is ready for transmission and presentation to the customer. This document may include additional explanations, illustrations, alternative configurations and options, as well as appendices and disclaimers. The quote document sent to the customer will typically also be stored in the CPQ, CRM or ERP system and/or archived in a document management system.

The *Service Delivery* phase starts when a settlement between the service provider and the customer is achieved. At this point, responsibilities and information are typically transferred from sales to other operational units within the organization. Although this phase typically is much longer and more intensive compared to the previous three phases, the details of service delivery are not in the focus of the Modularity Canvas. This field is mainly intended to capture in how far specifications and documents from sales determine the service delivery process and in how far these are revisited, e.g., for monitoring and controlling purposes, or for learning from service delivery for future sales activities or revising the service catalog.

Finally, the *Customer* field is in the center of the Modularity Canvas, as we consider customers to have an influence on all the surrounding fields. We expect the organization to make strategic decisions on service variety and modularity with their target customers in mind. Moreover, customers may also provide impulses to overthink and advance existing resource configurations, process definitions, service offerings, as well as sales channels and communications. Obviously, they are also involved in the operational processes in sales from request to quote, and, finally, as an external factor during service delivery.

5 Demonstration

For demonstrating the utility of the Modularity Canvas, we conducted workshops with five German industrial service providers from contract logistics, wind energy and automotive engineering sectors who were interested in advancing their service architecture and sales processes (Table 2). In order to incorporate different views on variety management and the quotation process, the workshops involved two to five members from various departments of each company (e.g., CEO, service portfolio managers, sales managers, IT system architects).

Table 2. Companies at which demonstration workshops took place

ID	*Industry*	*Offered services*	*Firm size (employees)*	*Revenue (million €)*
L1	Logistics	Packaging logistics, contract logistics, project business	200	18
L2	Logistics	Sea, air, and ground transportation; packaging, warehouse management	50	20
W1	Wind energy	Manufacturing of wind turbines, maintenance of onshore and offshore power plants, grid development	8000	5500
W2	Wind energy	Personnel leasing, technical services, technical training	400	5
A1	Auto-motive	Engineering services and consulting projects in the automotive industry	150	20

The half-day-workshops were conducted on-site of the respective company following a pre-defined agenda. Before discussing and filling out the actual Modularity Canvas, the agenda also included a first capturing of objectives and challenges in variety management and the quotation process. While the company members provided input to each of the Modularity Canvas fields (see exemplary notes in the right column of

Table 4), we as researchers acted as moderators and recorders. Hence, we were responsible for steering the discussions along the different fields of the canvas and tracking time. Figure 3 gives some impressions of how the Modularity Canvases looked liked at the end of the demonstration workshops.

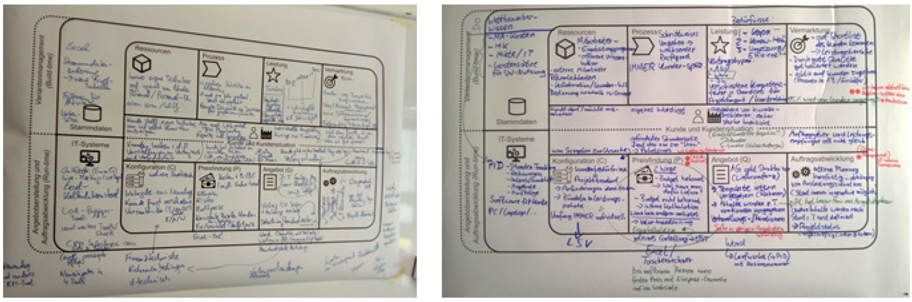


Figure 3. Impressions from the workshops with W1 (left) and A1 (right)

Table 3. Objectives and challenges of workshop companies

ID	*Objectives and challenges of workshop companies*
L1	• Price depends on the solvency of the individual customers • There is a strong motivation to introduce modular services in order to avoid to lag behind the already modular competitors • Knowledge needed for creating a quote exists solely in an employee's head
L2	• Prices are highly volatile (especially in sea transportation) • Completely different quotation processes per business unit • Time losses due to the back-and-forth communication with the customer
W1	• Business is strongly dependent on uncertainties in legislation • Service is offered according to the customer's technical knowledge • New services are identified in a provider-driven, not customer-driven, manner • Service packages are manually composed with no IT-support except for Excel files
W2	• Clear price structures with little room for negotiation • Pursue a clear expansion and internationalization strategy • First standardization approaches (e.g., unified quotation documents) • Rudimentary use of IT
A1	• Low number of highly powerful customers • Strong integration into the customers' development processes • Size of service offerings grow from single tasks to whole development projects • Tenders are fully adjusted to customer requirements, no service catalog with standardized services at the service provider

We conducted post-processing sessions after each workshop in order to discuss the contents of the Modularity Canvases and finally derive modularity potentials that we communicated to the participating companies together with a commented photo documentary of the workshops. We provided a summary of our observations relevant for developing a modularization vision and roadmap (see Table 3 for some excerpts for the five companies). We also explained our ideas for target visions and modularization potentials that we had derived from the status quo as captured in the Modularity Canvas (see Table 5 for an exemplary extract, here specifically for W1).

Table 4. Exemplary insights from the demonstration workshop at W1

Canvas field	*Guiding question*	*Exemplary notes in the Modularity Canvas from the demonstration workshop at W1*
Market	How are different customer segments addressed with service variants?	No differentiation or segmentation, but very different customers ranging from farmers to large-scale investors.
Outcome	Which service variants exist?	Three different packages (from basic to full-service), which can be fully customized.
Process	Which process variants exist?	Planned yearly servicing, unplanned incidents.
Resource	Which different resources are used?	External resources: cranes and ships.
Master Data	How are the variants represented in the master data?	Complex Microsoft Excel files. Changes in master data result in new, updated Excel files.
Customer	How does the customer influence the variance of marketing activities, outcome, processes and resources?	Customer provides own technicians (human resources) in order to absorb expertise.
	What interactions with the customer happen during the quotation process?	Iterative and long-lasting contract negotiations.
IT Systems	Which application systems support the quotation process?	Microsoft Outlook, Excel-based calculation and process tools, additional CRM system Salesforce.com, MS Sharepoint file sharing for projects.
Configure (C)	How is the service offering composed as a response to a request for a quote?	Initiated from wind turbine (physical product) sales, customer enquires about different service packages (e.g., different durations, basic vs. full-service).
Price (P)	How is the price of a bundle of service modules determined?	Cost plus calculation (not value-based), consideration of error rates, risks, and market prices.
Quote (Q)	How are quotation and contract documents designed?	Initial quote draft of 3-5 pages, later on: 50 pages contract plus 100 pages appendices.
Service Delivery	To what extent is information exchanged between sales and service delivery?	Defined handover-process to order processing department (offices next door) and project reviews.

We gathered qualitative feedback from the participants concerning the workshops and the Modularity Canvas as a central tool directly at the end of each workshop. In follow-up phone calls, we further asked them for feedback concerning the documentaries provided, including our ideas and recommendations. This feedback was generally positive as the participants appreciated the workshops, which in many cases offered them a new perspective on their own organization; and they also valued the potentials that we identified and explained.

The workshops and post-processing meetings allowed us to reflect on the Modularity Canvas. While the capturing of the status quo with the Modularity Canvas in the workshops seemed to be efficient, we realized that the derivation of the individualized potentials for modularity and the formulation of an overarching modularization strategy is not straightforward and requires a more systematic approach in addition to the canvas, which is currently lacking. Consequently, a design iteration that complements the canvas with a method (or guidelines at least) for more systematically deriving modularity potentials needs to be initialized before the artifact can be evaluated in a broader context.

Table 5. Identified modularity potentials at company W1

Target vision per layer	*Modularity potential*	*Potential actions*
Variety management: "Module catalog covers all customer demands"	• Launch of a modular service catalog • Communication of the modular service catalog to customers	• IT-supported modeling of the module catalog for integration into ERP system • Launch of online configurator
Quotation process: "Accelerated process with integrated IT support"	• Acceleration of processes • Increased understanding of customers • Internal selection support (click & choose) • Reduction of data input into different tools	• Integration of CPQ tool into the ERP system • Simplification of quotation documents

6 Conclusion and Outlook

With the Modularity Canvas, we present a novel artifact that can help organizations create visual models of their status quo in service modularity and identify improvement potentials. It is intended to support the initial phase of modularization initiatives in organizations when the status quo of variety management and quotation processes are captured and analyzed before the actual service modules are defined [7].

The motivation for this research was the limited adoption of the service modularity concept and corresponding methods in practice although a pre-study with experts pointed us to issues with variety management and quotation processes [27]. As it is of a similar complexity like the BMC, we expect our canvas to be a useful and appropriate tool for practice. From an academic perspective, this research also suggests a novel ontology of service modularity that includes both the strategic level of variety management and the operational level of the quotation process. Hence, this ontology offers a new and more holistic perspective on the service modularity concept apart from modularization methods and high-level conceptualizations that have dominated recent academic discussions.

While the general feedback on the current version of the Modularity Canvas from the demonstration workshops was positive, we realized that the derivation of specific modularity potentials and improvement measures during post-processing was not easy. We therefore plan to complement the canvas with a set of guidelines for systematically deriving modularity potentials. Once this is achieved, we also plan to conduct the DSRM steps *5. Evaluation* and *6. Communication*. Following March and Smith [26], the evaluation of the canvas will be directed towards criteria like completeness, simplicity, understandability, ease of use, as well as its fidelity with real world phenomena. The evaluation will most likely take place in a field study [33] together with a consulting company. Using the customer network of this consulting company, the Modularity Canvas will be applied in further real-life business contexts. In such settings, the consulting company will take the role of the moderator and we as researchers will be able to observe the workshops and to record the results. We expect that, with the increasing sample size, further design iterations are possible, leading to potential adjustments of the artifact's structure and improving its usefulness. Once no further adjustments are observable, the artifact together with practical insights will be *communicated* to various stakeholders in academia and practice.

References

1. Luczak, H., Gill, C., Sander, B.: Architecture for service engineering—the design and development of industrial service work. In: Advances in services innovations, pp. 47–63. Springer, Heidelberg (2007)
2. Pekkarinen, S., Ulkuniemi, P.: Modularity in developing business services by platform approach. Int. J. Logist. Manag. 19, 84–103 (2008)
3. Baldwin, C.Y., Clark, K.B.: Design Rules: The power of modularity. MIT Press (2000)
4. Böhmann, T., Junginger, M., Krcmar, H.: Modular service architectures: A concept and method for engineering IT services. In: 36th Hawaii International Conference on System Sciences (HICSS). (2003)
5. Dörbecker, R., Böhmann, T.: The Concept and Effects of Service Modularity -- A Literature Review. 46th Hawaii International Conference on System Sciences (HICSS). (2013)
6. Müller, F., Lubarski, A.: Schools of thought in service modularity. In: European Conference on Information Systems (ECIS), Istanbul, Turkey (2016)
7. Poeppelbuss, J., Lubarski, A.: A Classification Framework for Service Modularization Methods. Enterp. Model. Inf. Syst. Archit. (EMISAJ). 13, 14–1 (2018)
8. Frandsen, T.: Evolution of modularity literature: A 25-year bibliometric analysis. Int. J. Oper. Prod. Manag. 37, 703–747 (2017)
9. Starr, M.K.: Modular production – A 45-year-old concept. Int. J. Oper. Prod. Manag. 30, 7–19 (2010)
10. Beverungen, D., Lüttenberg, H., Wolf, V.: Recombinant Service Systems Engineering. Bus. Inf. Syst. Eng. 60, 377–391 (2018)
11. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons, Hoboken, NJ (2010)
12. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. 24, 45–78 (2007)

13. Bask, A., Lipponen, M., Rajahonka, M., Tinnilä, M.: Framework for modularity and customization: service perspective. J. Bus. Ind. Mark. 26, 306–319 (2011)
14. Tuunanen, T., Bask, A., Merisalo-Rantanen, H.: Typology for Modular Service Design: Review of Literature. Int. J. Serv. Sci. Manag. Eng. Technol. 3, 99–112 (2012)
15. Carlborg, P., Kindström, D.: Service process modularization and modular strategies. J. Bus. Ind. Mark. 29, 313–323 (2014)
16. Böttcher, M., Klingner, S.: Providing a method for composing modular B2B services. J. Bus. Ind. Mark. 26, 320–331 (2011)
17. Rahikka, E., Ulkuniemi, P., Pekkarinen, S.: Developing the value perception of the business customer through service modularity. J. Bus. Ind. Mark. 26, 357–367 (2011).
18. Vargo, S.L., Lusch, R.F.: Evolving To A New Dominant Logic for Marketing. J. Mark. 36, 1–10 (2004)
19. Dörbecker, R., Tokar, O., Heddaeus, D., Böhmann, T.: Evaluation der Multiple Domain Matrix Methode zur Modularisierung von Dienstleistungen am Beispiel eines Versorgungsnetzwerks für psychische Erkrankungen. In: Multikonferenz Wirtschafts-informatik (MKWI). (2014)
20. Buchmann, R.A.: Modeling product-service systems for the internet of things: The comvantage method. In: Domain-Specific Conceptual Modeling. pp. 417–437. Springer (2016)
21. Dörbecker, R., Böhmann, T.: FAMouS – Framework for Architecting Modular Services. In: 36th International Conference on Information Systems (ICIS), Fort Worth. (2015)
22. Broekhuis, M., van Offenbeek, M., Eissens-van der Laan, M.: What professionals consider when designing a modular service architecture? Int. J. Oper. Prod. Manag. 37, 748–770 (2017)
23. Kim, W.C., Mauborgne, R.: Charting your company's future. Harv. Bus. Rev. 80, 76–85 (2002)
24. Zolnowski, A., Weiß, C., Böhmann, T.: Representing Service Business Models with the Service Business Model Canvas – The Case of a Mobile Payment Service in the Retail Industry. In: 47th Hawaii International Conference on System Sciences, pp. 718–727 (2014)
25. Osterwalder, A.: The business model ontology: A proposition in a design science approach. (2004)
26. Uschold, M., Gruninger, M.: Ontologies: Principles, methods and applications. Knowl. Eng. Rev. 11, 93–136 (1996)
27. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. 15, 251–266 (1995)
28. Lubarski, A., Pöppelbuß, J.: Vertrieb industrienaher Dienstleistungen – Erkenntnisse aus der Windenergie-und Logistikbranche. Arbeitsbericht der Juniorprofessur für Industrienahe Dienstleistungen, Universität Bremen, Bremen (2017)
29. Heskett, J.L.: Lessons in the service sector. Harv. Bus. Rev. 65, 118–126 (1987)
30. Becker, J., Beverungen, D.F., Knackstedt, R.: The challenge of conceptual modeling for product–service systems: status-quo and perspectives for reference models and modeling languages. Inf. Syst. E-Bus. Manag. 8, 33–66 (2010)
31. Bullinger, H.-J., Fähnrich, K.-P., Meiren, T.: Service engineering—methodical development of new service products. Int. J. Prod. Econ. 85, 275–287 (2003)
32. Gartner IT Glossary: Configure, Price, Quote (CPQ) Application Suites, http://www.gartner.com/it-glossary/configure-price-quote-cpq-application-suites/
33. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. 28, 75–105 (2004)

Towards a Conceptualization of Capabilities for Innovating Business Models in the Industrial Internet of Things

David Soto Setzke[1], Tom Rödel[1], Markus Böhm[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Munich, Germany
{setzke,tom.roedel,markus.boehm,krcmar}@in.tum.de

Abstract. The emergence of Internet of Things (IoT) technologies offers promising value potentials for industrial manufacturers based on the combination of smart products and data-driven services. At the same time, many incumbent firms experience a threat to their traditional value proposition and are challenged to innovate and reconfigure their existing business models. However, many of these traditional manufacturers lack or are unaware of the required capabilities for successfully reinventing their business model using IoT technologies. We therefore adopt the lens of dynamic and operational capabilities and conduct an empirical analysis of organizational capabilities required for successful IoT-enabled business model innovation (BMI). Through an exploratory, qualitative study based on interviews with decision makers in industrial manufacturing companies and experts in practice-oriented research institutions, we identify eleven distinct dynamic and operational capabilities. Our findings provide useful insights for research and practice and advance the understanding of enablers in IoT-enabled BMI.

Keywords: Digital Transformation, Industrial Internet of Things, Dynamic Capabilities, Operational Capabilities, Business Model Innovation

1 Introduction

In recent years, the Internet of Things (IoT) received enormous attention in academic literature as well as industry practice and still remains a promising research area [1]. The emergence of IoT technologies and their application in the industrial context, also known as the Industrial Internet of Things (IIoT), changes competitive dynamics by erupting traditional market boundaries between industrial manufacturers, software providers, and technology start-ups [2, 3]. Traditional manufacturers are challenged to generate new value propositions through data-based services and predictive solutions [4] which often requires adaptation of existing business models [5]. The German automotive supplier Bosch, for example, uses IoT technologies to enable customers of its fleet management system to identify potential problems in advance and to analyze the driving behavior of individuals [4]. However, such change brings along numerous challenges and has major implications for incumbent firms [3, 6]. While traditional

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

manufacturers possess critical industry knowledge, they most likely face substantial skill gaps when it comes to IoT and related business model innovation (BMI) [2, 6]. Besides a lack of technological expertise in areas such as IoT infrastructure, data analytics, and software engineering, industrial manufacturers are required to rethink existing business model components and to implement new approaches towards customer relationship management, sales, and collaboration with technology providers [3]. All in all, the IoT constitutes an exogenous technological change to which industrial manufacturers need to react by adapting their business model in order to capture the value potential and to secure future competitiveness [7].

Existing academic work on IoT-enabled BMI is still young and little is known about how the change in business models actually occurs. Most notably, there is a missing perspective on how to overcome the identified challenges and barriers of IoT-enabled BMI. In fact, based on our assessment, current literature fails to analyze enablers of IoT-enabled BMI and to conceptualize relevant organizational capabilities. There is thus a strong need to better understand the complex underlying processes and drivers of successful IoT-enabled BMI. Overall, existing research does not clarify the nature of required organizational capabilities for IoT-enabled BMI. In this paper, we present a conceptualization of eleven organizational capabilities that are required for IoT-enabled BMI. We identified these capabilities through an exploratory approach involving semi-structured interviews with decision makers in the German manufacturing industry and experts in practice-oriented research institutions. In the following, we introduce our understanding of IoT-enabled BMI and organizational capabilities that we applied in our exploratory research.

2 Theoretical Background

2.1 IoT-enabled Business Model Innovation

Despite a large body of research, existing theory still misses a common understanding about both business model (BM) and BMI [6, 8]. Therefore, it is essential to define both concepts in the context of our study. Business models are described as "mental models" [9] that represent the underlying architecture of a firm's overall business [10]. The concept focuses on the underlying organizational structures, processes, and resources that enable value creation [9] and defines "[…] the manner by which the enterprise delivers value to customers, entices customers to pay for value, and converts those payments to profit" [11]. According to Foss and Saebi [8] BMI encompasses "[…] designed, novel, nontrivial changes to the key elements of a firm's business model and/or the architecture linking these elements". Following Tesch, Brillinger and Bilgeri [2], in the context of our study this includes both "the 'modification, reconfiguration and extension […] of existing business models' (business model development) as well as the design of 'fundamentally new and sometimes disruptive' business models (business model design)". Furthermore, we refer to BMI using IoT technologies as IoT-enabled BMI.

Literature on IoT-enabled BMI can be grouped into three major research streams. The first stream focuses on the analysis of business model patterns and frameworks for the IoT and identifies new patterns such as remote usage or condition monitoring [1, 4, 12-14]. While many studies analyze the influence of IoT on specific business model components and describe underlying changes [12, 13], other studies do not focus on single organizations but take a broader view on the overall IoT ecosystem by analyzing the interaction and collaboration of different players [14]. Second, a group of studies analyzes the process of IoT-enabled BMI itself [2, 6]. For instance, Tesch, Brillinger and Bilgeri [2] apply a stage-gate model to IoT-enabled BMI and identify a semi-structured, iterative process. Moreover, current literature builds on processes identified in product development research, such as innovation stages in the process of IoT-enabled BMI [6]. Third, an emerging stream of literature analyzes challenges and barriers in IoT-enabled BMI [4, 6, 15]. Thereby, challenges are analyzed from both a technical and business perspective [6]. Manufacturers require new capabilities to incorporate software, data analytics, and data-based service offerings [2, 15]. All in all, companies need to develop capabilities to master both technology and business-related challenges in order to successfully implement IoT-enabled BMI [4]. However, current research is missing a close analysis of such organizational capabilities.

2.2 Organizational Capabilities

In this paper, we conceptualize organizational capabilities as dynamic and operational capabilities. The concept of dynamic capabilities was first introduced to better address the characteristics of today's volatile business environments and markets [16, 17]. They are described as "higher-order organizational capabilities" [18] that enable incumbent firms to modify existing capabilities, organizational structures, and even company culture [7, 18, 19]. The framework refined by Teece [20] distinguishes three basic dimensions of dynamic capabilities and differentiates the underlying organizational processes into the classes of sensing, seizing, and reconfiguration. Sensing capabilities encompass the organizational ability to discover opportunities related to technological developments as well as changes in customer requirements and the overall market [20, 21]. Seizing capabilities mainly encompass processes related to organizational value generation as well as new product development or service innovation [20]. Reconfiguration capabilities are based on processes for the alignment and realignment of organizational assets in order to meet new requirements [20]. These capabilities can address organizational topics such as decentralization or co-specialization and encompass critical processes of organizational knowledge management [20, 21].

Existing literature argues for the need to differentiate between different levels of hierarchy of organizational capabilities in order to reduce confusion about the concept and to eliminate its "tautological feel" [22]. Therefore, we distinguish two main classes of organizational capabilities: Dynamic capabilities and operational capabilities [23, 24]. Operational capabilities, also described as ordinary [16] or zero-level capabilities [17], encompass the operational function of a firm and enable the value proposition of a business model [22]. They are responsible for the execution of daily business operations and can be described as "how you earn your living" capabilities [17, 22]. In contrast, dynamic capabilities represent “how you change your operational routines” capabilities [22].

2.3 Dynamic Capabilities as Antecedents of Business Model Innovation

Several scholars regard dynamic capabilities as internal antecedents and drivers of BMI processes [8, 25]. Dynamic capabilities are integral to BMI as they enable firms to design and implement effective new business models [20, 25]. In addition, BMI requires strong dynamic capabilities as it involves a complex process of organizational and strategic renewal [19]. Besides strong sensing capabilities to realize the need for change, seizing capabilities are required for the modification and redesign of existing business models [19]. However, Leih, Linden and Teece [19] argue that capabilities for organizational reconfiguration and actual implementation of the business model are most critical, as BMI processes affect organizational boundaries, internal structures, and even company culture. Several authors build on the dynamic capabilities framework to advance theory on enabling capabilities. Mezger [18] conceptualizes BMI itself as a “distinct dynamic capability” and identifies corresponding organizational routines and processes. He uses the original framework by Teece [20] to disaggregate BMI dynamic capability into the dimensions of sensing, seizing, and reconfiguring capabilities. Thereby, “business model sensing” capabilities enable opportunity recognition by monitoring competition, market developments, and changes in industry-wide business models [18]. “Technology sensing” capabilities allow for a systematic assessment of technological possibilities and the exploration of new ideas. Seizing capabilities comprise innovation activities for the design and configuration of business models. Actual business model implementation is realized by reconfiguring capabilities that facilitate the realignment of operational capabilities and resources [18].

3 Methodology

We apply an exploratory, qualitative research design based on interviews with knowledgeable experts from the field to explore and describe the phenomenon of IoT-enabled BMI. We argue that the complex and highly context-specific nature of organizational capabilities is well-suited for the use of qualitative research methods. This approach allows us to generate rich theoretical insights from complex organizational decisions and processes. Further, the present study draws on evidence from multiple organizations to include several perspectives on the researched phenomenon. In the following, we describe our approaches for data collection and analysis in more detail.

3.1 Empirical setting

Regarding our industry interviews, we apply an industry focus on German small and middle sized enterprises (SMEs) in machinery and plant engineering to control for industry, regional, and strategic context [18]. The German industry is characterized by many highly specialized SMEs that contribute large economic value. Although many of the firms are global market leaders in specific segments, their positions are threatened by ongoing commoditization of machinery and by new competition arising from outside of the traditional manufacturing industry [26]. Thereby, most SMEs in machinery and plant engineering represent typical product-oriented manufacturers that are now challenged to innovate their business models [13, 26]. In addition, SMEs are likely to possess fewer resources as compared to industrial giants such as GE. Thus, they might lack sufficient capacities to react to technological change appropriately. The European Commission defines SMEs based on staff headcount and either turnover, or balance sheet total [27]. Thereby, a company qualifies as SME if it does not have more than 249 employees and its annual turnover does not exceed 50 Million €. However, many firms of the so-called "Mittelstand" in German machinery and plant engineering do not meet these requirements. Therefore, we apply the broader definition of SMEs provided by the Institute for SME research in Bonn to our company sample. Consequently, we also consider companies where the majority of company shares is hold by up to two natural person or their family members, given that these shareholders are active in the executive board [28].

We use theoretical sampling [29] to identify appropriate organizations for the empirical analysis. The objective of the selection process was to identify SMEs in the industry that already engage in IoT-enabled BMI and that experience the related transformation towards product-service combinations. We conducted an online search, using information from industry association websites and trade journals, to identify promising manufacturers for our research approach. We then gathered more specific information on single companies based on their corporate websites, product and service portfolios, and related press articles. In total, we contacted 50 individuals of 37 different companies, from which 17 executives replied. Some of them declined participation due to reasons of confidentiality, time pressure, or lack of experience. Eventually, we were able to schedule interviews with representatives from seven different SMEs. Our sample comprises six machine manufacturers and one electrical component supplier. All SMEs are headquartered in Germany but are present on international markets and often conduct global operations.

3.2 Data collection and analysis

In total, we conducted eight qualitative interviews with industry experts on IoT-enabled BMI. Seven interviews represent conversations with representatives of manufacturing firms. Thereby, we performed one interview per organization with each one executive. Moreover, we conducted one additional interview with an industry expert from a renowned research institution at the beginning of the data collection process. The interview was not firm-specific and rather explorative. We used the insights to generate a first understanding of IoT-enabled BMI in machinery and plant engineering and to further refine our interview guideline. **Table 1** represents an overview of all conducted interviews and the respective interview partners. Thereby, all interviewees were required to have at least three years of industry or research experience and, in the case of manufacturing organizations, to hold a managing position, preferably senior management, in research and development, business development, or product and innovation management.

Table 1. Overview of interviewed experts (M = manufacturing organization; R = research institution)

ID	*Expert role*	*Business sector*	*Founding year*	*Number of employees*	*Sales turnover*
M1	Head of Business Development	Packaging machinery and solutions	1869	2.500	€ 835 Million (2017)
M2	Senior Business Development Manager	Packaging machinery and solutions	1922	2.250	€ 350 Million (2017)
M3	Head of Product Engineering	Raw material processing and recycling machinery	1969	400	€ 100 Million (2017)
M4	Chief Information Officer	Environmental simulation and welding machinery	1913	8.200	€ 1,2 Billion (2017)
M5	Head of Digitalization	Packaging machinery and solutions	1961	5.065	€ 1 Billion (2017)
M6	Head of Machinery Solutions	Electrical component supplier	1850	4.700	€ 740 Million (2017)
M7	Head of Process Engineering	Water processing and machinery	1989	220	€ 19 Million (2016)
R1	Research Expert on Digital BMI	Research institution	1995	25.000	n/a

The interviews were recorded and transcribed afterwards. We used Qualitative Content Analysis as introduced by Mayring [30] to evaluate the transcribed expert interviews. While the initial categories were derived directly from the text basis using an open coding approach, we developed the main categories in close relation with existing theory on organizational BMI capabilities [18]. Challenges encountered by the organization on their way to IoT-enabled BMI constitute the basis of our category system. Thereby, a challenge comprises a situation that is described as being problematic and relatively new to the firm. Moreover, it cannot be solved with existing organizational processes, but requires management attention and dedicated investments. In addition, the challenge must not be firm-specific but can be transferred to the context of other organizations. The coding itself was conducted separately for each case study in order to allow for within-case analysis before aggregating the results. We then used existing literature on organizational capabilities to develop main categories for the identified challenges. The main categories group similar findings and allow us to identify critical capabilities for IoT-enabled BMI.

4 Organizational Capabilities for IoT-enabled BMI

We propose a conceptualization of IoT-enabled BMI organizational capabilities to link our findings to extant literature. We apply the lens of dynamic and operational

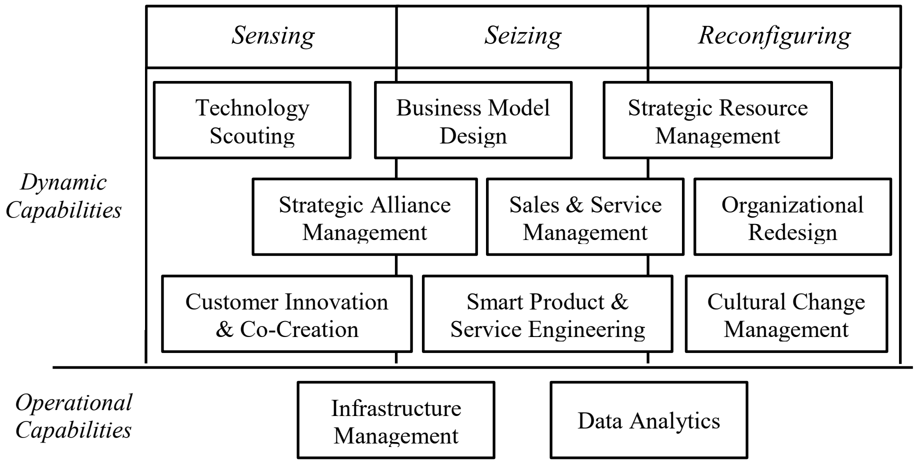


Figure 1. Conceptualized organizational capabilities for IoT-enabled BMI

capabilities to interpret our findings and group them according to the three dimensions of sensing, seizing, and reconfiguring dynamic capabilities [20]. Moreover, we use the concepts of dynamic and operational capabilities to distinguish between different types of organizational capabilities and the level of hierarchy on which they operate. Figure 1 presents our theoretical model that integrates the empirical findings into existing theory on BMI capabilities. We do not interpret the identified dynamic capabilities as purely sensing, seizing, or reconfiguring since they are often based on intertwined processes that relate to more than one capability dimension. Therefore, we interpret the three dimensions rather as a continuum and allocate identified dynamic capabilities in accordance to their main function and purpose. Furthermore, the model does not imply a strict chronological order. Although sensing capabilities are clearly needed at the beginning of the innovation process, the process of BMI is of iterative nature [2].

(1) Technology Scouting: A key challenge described by interviewees from all organizations in our sample is the understanding of IoT as a technology itself. Moreover, companies need to track the trends in technology development and assess the potentials of current IoT technologies. They first need to identify and then test appropriate solutions for the implementation within the own business environment:

> "*To a certain degree we are confronted with a real flood of suppliers. […] Consequently, there are incredibly many service providers and suppliers of IoT technologies that are entering the market. And […] it is a big challenge to […] identify the right technologies that are appropriate for the own use case.*" (M5)

This underlines that without a critical assessment at the beginning of the BMI process, companies will not be able to fully leverage the potential of IoT technologies and establish them at the foundation of their new business models. The capability "technology scouting" guides the evaluation process and increase the overall understanding of the technology itself.

(2) Infrastructure Management: Another challenge is the establishment of infrastructure that enables interconnection. Manufacturers need to install the required sensor technology on the machinery and establish network connections. Thereby, data and network security are highly important and need to be assured at all time:

"*Usually, our clients have their internal networks which are secured and protected. This is a major topic nowadays. Network security. But you have to access these networks. You have to access the client's network from the outside to do your job and this a major technical challenge"* (M7)

This also includes important decisions with regards to infrastructure for data storage, data processing, and data utilization. Many SMEs in machinery and plant engineering have no or little experience when it comes to sensor technology and IT security. Therefore, Infrastructure Management represents a critical IoT-enabled BMI capability. It encompasses the ability to establish and manage the required IoT-infrastructure for data generation and data-based value creation.

(3) Data Analytics is another organizational capability that is required to address the challenge of IoT technology as an enabler of BMI. It constitutes the capability to generate customer value from machine and process data, and to develop related software applications for data-based services. Therefore, organizations need to expand their existing skills in software engineering and build up critical expertise in areas such as big data or data science:

"I believe that one challenge that many companies face is to extensively collect data, to retrieve this data, to analyze it, and to draw the right conclusions in order to generate value for customers and for themselves." (M3)

(4) Business Model Design: Besides technology-related capabilities to implement IoT technologies as the necessary foundation, actual business model design is a key challenge. Organizations need to map business opportunities and define the corresponding use cases. This includes the design of new value propositions to meet emerging customer demands and to clearly segment existing and potentially new customer groups. Altogether, business model design depends on entrepreneurial processes which enable the exploration of new value propositions. Key decision makers need to promote the idea of recurring revenues and design appropriate revenue models. We therefore propose the organizational capability of Business Model Design that enables the organization to identify IoT-enabled value propositions and to design the corresponding BM. The capability is based on a systematic process for the exploration of new value opportunities and use cases. It is required to challenge the existing

business models and to implement a systematic and strategic approach towards business model design:

"*Well, everyone has already heard at one point about leasing or predictive maintenance. But to systematically list 80 different business models and to analyze what fits to our company, that has not happened in the beginning. It was all very casual and rather informal"* (M1)

(5) Strategic Resource Management: The implementation of IoT-enabled BMI and the development of related organizational capabilities depends largely on the right resource endowments. Companies that engage in IoT-enabled BMI need to identify critical know-how and develop it within the organization:

"*I also think that we should develop a lot of these competencies internally and not source them from the outside. Because at the moment it is quite difficult to foresee which competencies will be most critical for our future business."* (M6)

This also emphasizes the need for qualified employees. Many organizations are highly dependent on specialists that bring required know-how into the organization. Several companies in our sample have mentioned challenges with regards to the location of their headquarters that are often situated in rural areas. Besides the lack of know-how, the allocation of resources to innovation-related activities in addition to the current operations represents a key challenge, especially because most of the companies from our sample face exceptional good order positions and are working at full capacity. Our proposed capability allows companies to manage internal competition for resources and to pursue BMI activities without affecting ongoing operations negatively. Moreover, it encompasses the ability to identify areas of expertise that are best developed internally in order to gain competitive advantage in the long run.

(6) Customer Innovation & Co-Creation: Customer relations represent another main challenge faced by our sample organizations. On the one hand, industrial manufacturers require a certain level of openness to collaborate with customers and consider their input for product and service innovation. They need to understand the value of such co-innovation and establish the processes for collaborative innovation. However, this often contradicts the traditional mindset of SMEs in machinery and plant engineering. Many organizations have been very critical towards open innovation in the past and now face difficulties to open themselves and promote a new understanding of their clients as valuable business partners:

"We agree that it is important to understand customers more as partners. In my opinion that is inevitable for the survival in global competition." (M3)

(7) Sales & Service Management: Our interviewees have pointed out the necessity to adapt existing marketing and sales processes. They need to create new ways on how to approach the client in order to demonstrate the value of data-based IoT services. The responsible sales teams need to understand the business value arising from software

applications as well as smart services and integrate the idea of recurring revenues in contrast to onetime sales. They also have to convince clients of the new value proposition and overcome customer concerns with regards to data privacy:

"Consequently, we have to change the way we approach our clients and how we are selling our solutions. So far, our machinery has never been online. We sold pure offline machinery that is usually located at […] storages at client site […]. This has never been an issue for them. Actually, they are very sensitive when it comes to external data and network connections, especially data sharing." (M5)

Moreover, the sales system needs to internalize a new understanding of services and digital products. In accordance with new revenue models, a shift from a product-centric towards a service-centric sales system might be required. We propose IoT Sales & Service Management as another capability to address the challenges at the front end of the IoT business model. This capability allows to market IoT-enabled products and services appropriately by reconfiguring established sales processes and by designing appropriate IoT sales and service strategies.

(8) Strategic Alliance Management: Many traditional manufacturers in machinery and plant engineering have only recently started to engage in open discussions on market and technological developments. In fact, some of them have never built on external solutions before to realize their product offerings. That is why they need to promote an integral organizational openness towards external collaboration:

"I am convinced that only those companies will succeed in IoT-enabled BMI that engage in strategic alliances. This means to cooperate with others along the value chain, with regards to data usage and data processing, if necessary with competitors [...]. Only if these networks are created, which by the way is totally untypical for German machinery and plant engineering, […] success […] will be possible." (M1)

We propose that organizations need to develop Strategic Alliance Management capabilities to collaborate with external partners and networks in order to complement their existing capabilities. Moreover, they need to establish organizational processes that help to identify potential partners and to build up strategic alliances.

(9) Smart Product & Service Engineering: The value creation itself represents another major challenge. IoT-enabled BMI affects existing product innovation processes that so far are mainly oriented towards the development of physical products and add-on services such as repair and maintenance. Established product engineering processes are often not appropriate for the development of smart, digital products. Moreover, companies need to develop and implement a new understanding of a value creation that is based on machine and process data:

"Many organizations have no experience when it comes to data-based services. […] For example, if you do not sell machinery anymore but provide operator models you have to abandon the idea of onetime sales and implement processes for lifecycle-

services and recurring revenues. But this requires a huge shift in mindset with regards to value creation." (R1)

Therefore, we propose Smart Product & Service Engineering as an essential organizational capability for IoT-enabled BMI. It enables the organization to redesign existing product and service engineering processes and to develop smart products and smart services for data-based value creation.

(10) Organizational Redesign: Both scope and complexity of the organizational implementation of IoT-enabled BMI represent major challenges for our sample companies. Besides the necessity to redesign many critical organizational processes, nearly all organizational departments are affected by business model change. This emphasizes the need for a comprehensive transformation process that incorporates all organizational departments. Such complexity of implementation likely overwhelms traditional industrial manufactures:

"Another point is that we realized that the whole topic around digitalization, transformation, changing market requirements, and organizational culture involves such high complexity that we feel overwhelmed and that very likely we are not able to cope with this transformation on our own, organically." (M2)

Although all of the identified organizational capabilities enable organizations to reconfigure organizational processes, we propose a distinct capability of Organizational Redesign that allows to reconfigure organizational structures and support processes as well as to reallocate responsibilities to organizational units.

(11) Cultural Change Management: Many of the above-mentioned challenges and capabilities already point out the importance of a change in organizational mindset. Thereby, organizations not only need to challenge their existing business models, but also need to realize the importance of change in the first place. Despite current favorable market conditions, they need to take notice of the developments in the industry and raise overall awareness and openness towards change:

"It is also a very comfortable position to just say and acknowledge something could happen. I mean our order books are so full and the situation at the moment is just heavenly." (M4)

Organizations need to develop a certain organizational mindset that allows them to observe changes in market and technology and to initiate first actions. We therefore propose Cultural Change Management as an organizational capability that enables manufacturers to induce and manage cultural change throughout the organization. Thereby, it promotes a culture that values exploration and raises the openness towards IoT-enabled BMI.

5 Discussion, Contributions, and Limitations

The findings from our qualitative study support our understanding of dynamic capabilities derived from existing literature. They encompass a collective activity that enables organizations to systematically modify its operating routines [31]. We also find evidence for the key role of top management in the reconfiguration process [32]. Moreover, the empirical findings show the importance of sensing, seizing, and reconfiguring dynamic capabilities for IoT-enabled BMI [20]. In order to cope with technological change such as the emergence of IoT technologies, organizations need to reconfigure their existing resources as well as operational capabilities and establish new organizational processes [7]. Several of our identified capabilities could also be applied to general BMI (e.g., Business Model Design or Technology Scouting) or to data-driven BMI (e.g. Data Analytics), i.e., BMI based purely on the use of data analytics. However, capabilities such as Smart Product & Service Engineering go beyond the mere collection and analysis of data. While data-driven business models focus on "acquisition of data, its subsequent aggregation, the analysis of data […], and actions that are triggered" [33], we argue that IoT-enabled business models can be interpreted as an instance of data-driven business models that focus on more specific aspects such as enriching physical products with digital services [1]. However, future research could further explore how IoT-enabled BMI differs from more general data-driven BMI.

Our study contributes to literature on dynamic capabilities to advance theory on enabling factors in BMI [8]. Thereby, our set of organizational capabilities confirms the relevance of previously identified dimensions of dynamic capabilities in BMI research. Furthermore, we reduce the abstractness of the dynamic capabilities framework [5] by analyzing the underlying processes and providing a conceptualization of concrete capabilities.

The proposed findings have several important implications for industry practice and managerial decisions. In essence, SMEs are required to undertake a systematic assessment of their existing organizational capabilities and to define a set of capabilities required for their individual BMI aspirations. Key decision makers in the organization need to realize the need for change and interpret the value opportunities of IoT technology accordingly. IoT-enabled BMI very likely affects the entire organization and requires organizational redesign and restructuring. One key insight for managers is the necessity of cultural change. Leadership needs to promote an overall organizational openness towards external exploration and to overcome traditional thinking. Overall, we believe that our conceptualization of capabilities assists practicing managers in making informed decisions about the required investments in capability development and in reflecting on IoT-enabled BMI in general. Thereby, the practical implications are not limited to SMEs in machinery and plant engineering.

Our findings are not free from limitations. First, our model does not represent a complete set of capabilities and several capabilities might overlap to some degree or depend on each other (for example Organizational Redesign and Cultural Change Management). Organizational capabilities are highly context-dependent, and every incumbent firm faces different capability endowments [16]. Therefore, there is no definite set of key capabilities and our findings need to be interpreted within the given organizational context of a firm. Although we propose that our proposed capabilities lead to successful IoT-enabled BMI, we do not measure the interrelation with firm performance nor do we provide any evidence of a positive effect of the realization of our capabilities on actual BMI implementation. In fact, we argue that a capability-based conceptualization of IoT-enabled BMI alone cannot explain successful IoT-enabled BMI and superior performance since many factors need to be taken into account for an analysis of firm performance [16]. Furthermore, while exploratory, qualitative research approaches offer great potential to add new perspectives and extend existing theory, our relatively small expert sample limits the generalizability of the findings [29].

As mentioned in section 3.1, our empirical settings is focused on German SMEs in machinery and plant engineering in order to control for industry, regional, and strategic context. Furthermore, we believe that these SMEs are, due to their limited resources, under high pressure to build up relevant capabilities and thus represent an interesting context for our study. On the other hand, our identified capabilities could also be specific to SMEs in our chosen context while capabilities for large corporations or companies in other regions could be different. Strategic Alliance Management, for example, could be less critical for large corporations due to their extensive sets of existing resources. We regard our results as a first step towards an exhaustive conceptualization of capabilities for IoT-enabled BMI and invite other researchers to verify, extend, or adjust our set of capabilities by replicating our study in different contexts.

6 Conclusion and Opportunities for Future Research

The emergence of IoT technologies brings along new business opportunities in industrial manufacturing. However, IoT-enabled BMI constitutes a highly complex transformation process and implicates severe challenges [6]. Thus, the main purpose of this paper is to advance research on organizational capabilities that are required to master the challenges of IoT-enabled BMI. We identify several dynamic and operational capabilities that represent enablers of IoT-enabled BMI. Overall, organizations are required to assess their existing capability endowment and strategically invest in IoT-enabled BMI capabilities to seize the value opportunities of the Internet of Things in industrial manufacturing. Thereby, our empirical findings contribute to understanding key enablers and antecedents in BMI [25]. Finally, they outline a promising field for future research on IoT-enabled BMI.

IoT-enabled BMI in industrial manufacturing offers a promising area for future research in both information systems and strategic management literature. Especially the concepts of BMI and dynamic capabilities require additional empirical studies to advance existing conceptualization and overall understanding. While we present a rather aggregated view on different capabilities, future studies could focus on distinct capabilities and analyze underlying processes and resources in detail.

Furthermore, future research could include large-scale, empirical studies with longitudinal design. Such studies would allow observing the entire process of IoT-enabled BMI and could provide important insights on performance outcomes and the interrelation of different organizational capabilities. In addition, studies applying a retrospective analysis on success cases could provide interesting benchmarks and contribute to a comprehensive understanding of IoT-enabled BMI.

References

1. Fleisch, E., Weinberger, M., Wortmann, F.: Geschäftsmodelle im Internet der Dinge. HMD Praxis der Wirtschaftsinformatik 51, 812-826 (2014)
2. Tesch, J.F., Brillinger, A.-S., Bilgeri, D.: Internet of Things Business Model Innovation and the Stage-Gate Process: An Exploratory Analysis. International Journal of Innovation Management 21, (2017)
3. Bilgeri, D., Wortmann, F., Fleisch, E.: How Digital Transformation Affects Large Manufacturing Companies' Organization. 38th International Conference on Information Systems, Seoul, South Korea (2017)
4. Weinberger, M., Bilgeri, D., Fleisch, E.: IoT business models in an industrial context. at - Automatisierungstechnik 64, (2016)
5. Achtenhagen, L., Melin, L., Naldi, L.: Dynamics of Business Models – Strategizing, Critical Capabilities and Activities for Sustained Value Creation. Long Range Planning 46, 427-442 (2013)
6. Bilgeri, D., Wortmann, F.: Barriers to IoT business model innovation. 13. Internationale Tagung Wirtschaftsinformatik, St. Gallen, Switzerland (2017)
7. Lavie, D.: Capability Reconfiguration: An Analysis Of Incumbent Responses To Technological Change. Academy of Management Review 31, 153-174 (2006)
8. Foss, N.J., Saebi, T.: Fifteen Years of Research on Business Model Innovation. Journal of Management 43, 200-227 (2017)
9. Foss, N.J., Saebi, T.: Business Models and Business Model Innovation: Bringing Organization into the Discussion. In: Foss, N.J., Saebi, T. (eds.) Business Model Innovation: The Organizational Dimension. Oxford University Press, Oxford, UK (2015)
10. Teece, D.J.: Business models and dynamic capabilities. Long Range Planning 51, 40-49 (2018)
11. Teece, D.J.: Business Models, Business Strategy and Innovation. Long Range Planning 43, 172-194 (2010)
12. Dijkman, R.M., Sprenkels, B., Peeters, T., Janssen, A.: Business models for the Internet of Things. International Journal of Information Management 35, 672-678 (2015)
13. Laudien, S.M., Daxböck, B.: The Influence of the Industrial Internet of Things on Business Model Design: A Qualitative-Empirical Analysis. International Journal of Innovation Management 20, 16-40 (2016)
14. Saarikko, T., Westergren, U.H., Blomquist, T.: The Internet of Things: Are you ready for what's coming? Business Horizons 60, 667-676 (2017)
15. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming companies. Harvard Business Review 93, 96-114 (2015)
16. Teece, D.J., Pisano, G.P., Shuen, A.: Dynamic Capabilities and Strategic Management. Strategic Management Journal 18, 509-533 (1997)
17. Winter, S.G.: Understanding dynamic capabilities. Strategic Management Journal 24, 991-995 (2003)
18. Mezger, F.: Toward a capability-based conceptualization of business model innovation: insights from an explorative study. R&D Management 44, 429-449 (2014)

19. Leih, S., Linden, G., Teece, D.J.: Business Model Innovation and Organizational Design: A Dynamic Capabilities Perspective. In: Foss, N.J., Saebi, T. (eds.) Business Model Innovation: The Organizational Dimension. Oxford University Press, Oxford (2015)
20. Teece, D.J.: Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal 28, 1319-1350 (2007)
21. Eriksson, T.: Processes, antecedents and outcomes of dynamic capabilities. Scandinavian Journal of Management 30, 65-82 (2014)
22. Cepeda, G., Vera, D.: Dynamic capabilities and operational capabilities: A knowledge management perspective. Journal of Business Research 60, 426-437 (2007)
23. Helfat, C.E., Peteraf, M.A.: The dynamic resource-based view: capability lifecycles. Strategic Management Journal 24, 997-1010 (2003)
24. Hoberg, P., Krcmar, H.: Vendor Capabilities in ITO Research: A Systematization and Critical Review. In: Oshri, I., Kotlarsky, J., L.P., W. (eds.) Advances in Global Sourcing. Models, Governance, and Relationships. Global Sourcing 2013, vol. 163. Springer, Berlin, Heidelberg, Berlin, Heidelberg (2013)
25. Bouncken, R.B., Lehmann, C., Fellnhofer, K.: The role of entrepreneurial orientation and modularity for business model innovation in service companies. International Journal of Entrepreneurial Venturing 8, (2016)
26. Emmrich, V., Döbele, M., Bauernhansl, T., Paulus-Rohmer, D., Schatz, A., Weskamp, M.: Geschäftsmodell-Innovation durch Industrie 4.0: Chancen und Risiken für den Maschinen- und Anlagenbau. Fraunhofer IPA / Dr. Wieselhuber & Partner GmbH (2015)
27. European Comission: What is an SME? http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en (Accessed: 08/09/2018)
28. IfM Bonn: Mittelstandsdefinition des IfM Bonn. https://www.ifm-bonn.org/definitionen/mittelstandsdefinition-des-ifm-bonn/ (Accessed: 14/09/2018)
29. Eisenhardt, K.M.: Building Theories from Case Study Research. Academy of Management Review 14, 532-550 (1989)
30. Mayring, P.: Qualitative Inhaltsanalyse. In: Flick, U., Kardoff, E.v., Keupp, H., Rosensteil, L.v., Wolff, S. (eds.) Handbuch qualitative Forschung : Grundlagen, Konzepte, Methoden und Anwendungen. Beltz - Psychologie Verl. Union, Munich, Germany (1991)
31. Zollo, M., Winter, S.G.: Deliberate Learning and the Evolution of Dynamic Capabilities. Organization Science 13, 339-351 (2002)
32. Zahra, S.A., Sapienza, H.J., Davidsson, P.: Entrepreneurship and Dynamic Capabilities: A Review, Model and Research Agenda. Journal of Management Studies 43, 917-955 (2006)
33. Zolnowski, A., Christiansen, T.G., Jan: Business Model Transformation Patterns of Data-Driven Innovations. 24th European Conference on Information Systems, İstanbul,Turkey (2016)

A Taxonomy of Barriers to Digital Transformation

Kristin Vogelsang[1], Kirsten Liere-Netheler[1], Sven Packmohr[2,3], and Uwe Hoppe[1]

[1] Universität Osnabrück, IMU, Osnabrück, Germany
{kristin.vogelsang, kirsten.liere-netheler, uwe.hoppe}@uos.de
[2] Malmö University, TS / DVMT, Malmö, Sweden
sven.packmohr@mau.se
[3] Bahçeşehir University, IISBF, Istanbul, Turkey
sven.packmohr@eas.bau.edu.tr

Abstract. Companies expect significant long-term gains in efficiency and productivity through digital transformation (DT). New ways of combining products, processes, and data-driven services, as well as new business models emerge. However, the rapid development of the DT leads to constraints regarding its realization. Barriers hinder companies to realize possible advantages out of DT. If firms promptly recognize potential barriers, they can reflect upon these challenges and can take well-coordinated countermeasures. Social, technical and socio-technical problems address different stakeholder and ask for specific solutions. Therefore, our study aims at developing a taxonomy for barriers to DT to enable researchers and practitioners to identify and classify existing barriers. For deriving the dimensions and characteristics, we collected data by conducting 46 semi-structured interviews with experts and enriched these by looking at the literature on DT barriers.

Keywords: Barriers, Digital Transformation, Taxonomy, Qualitative Research

1 Introduction

The increasing digitalization affects the industry and industrial processes [1]. A second machine age is announced to describe significant impacts of technology on the way of working [2]. Digital transformation (DT) is understood as the "use of new digital technologies (social media, mobile, analytics or embedded devices) to enable major business improvements (such as enhancing customer experience, streamlining operations or creating new business models)." [3] It is regarded as a major change in business and society [4] and is often described as an ongoing process [5]. One implication of digital transformation (DT) is the overall digitalization and cross-linking of the value creation process [6] as information and communication technologies merge with production processes. Digital transformation marks the whole business and must be taken up by the enterprises' strategy [7]. As DT affects people, processes, and products on all levels [2], it implies more than just an application of technology. Because of these overall implications, enterprises expect significant long-term gains in

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

efficiency and productivity by applying DT [4]. New ways of combining products and services [6] as well as new business models such as digital platforms emerge [8, 9].

Even though significant benefits are expected, many firms still struggle to realize transformation potential due to different barriers [10]. Only every sixth company in Germany has a high willingness for digital transformation [11]. So, the adoption of digital innovations is very faltering. Understanding present obstacles to planning countermeasures is critical. The rapid development of the digital transformation in companies leads to problems regarding the realization because of the development, diffusion, and implementation of new digitalized processes face many difficulties [12]. These difficulties or barriers (terms are used synonymously here) are understood as something that stands in the way of the realization of a project or the like and cannot be mastered without further ado.

The barriers to innovation approach [13] is often used to identify barriers and group these. Regarding DT, lists of barriers in specific technological contexts and different industries already exist [14, 15]. However, comparing and analyzing these to come to an overall view is still difficult. Taxonomies, which show dimensions and their characteristics, enable researchers and practitioners to understand, communicate and apply research findings in a more comprehensive way than lists. They help to "[…] structure or organize the body of knowledge" [16]. Describing taxonomies of barriers to DT are still missing. Thus, the goal of our study is to derive a taxonomy of DT barriers to contribute to the barriers literature. Different dimensions which can explain and classify these barriers are identified and analyzed. Categorizations help to generalize and lead to a better understanding. Therefore, the research question is: How can barriers to digital transformation be categorized into a taxonomy?

As the domain lacks theoretical foundation [17], we expect a qualitative approach to be useful to gain insights on barriers. On a broad database of 46 semi-structured interviews, we empirically derive the taxonomy in accordance with the procedure from Nickerson [18]. We use the collected material to develop a theoretical frame [19] which assists with overcoming the barriers. Results from the literature on DT barriers is also used to enrich the understanding.

For our study, we follow a holistic view. Digital transformation is said to transform whole value chains [6] which leads to a paradigm shift in business and focuses on the organization, rather than technological challenges [20]. Many researchers concentrate on distinct technologies or departments to describe the changes. As DT is still ongoing and many enterprises are at the beginning of the digital change, holistic approaches that regard the whole process of value creation are rare.

In the next chapter, we will define the term digital transformation and give an overview of existing classifications as well as barriers to DT literature. The methodology is presented in the third chapter. Afterward, the taxonomy is deduced from the interviews. Results from an evaluation and an outlook are shown. Finally, this paper highlights the contribution and describes limitations.

2 Related Work

As the field of DT is still rather new and many areas, that are touched by this, do exist, there is a certain need to classify and sort correlated phenomena. Several approaches exist, that set up classifications in distinct fields of DT research such as technologies (like cloud computing and big data) [21], digital strategies [17, 22, 23] as well as product and service enhancements [24–26]. Other researchers concentrate on the technology focus such as system architectures [27]. Furthermore, there is research on correlated topics like legal and regulatory frameworks [28] and DT research (methods) [29]. Taxonomies for certain industries also exist [30, 31]. These taxonomies help to structure the field of research on DT which is necessary because this is still a growing topic missing theorization [17]. Researchers already examined the positive aspects of DT success and developed a taxonomy of benefits [32]. We consider a systematic reappraisal of the opposite side of DT, which are the barriers that hinder a successful transformation, as a further valuable approach to understand DT. Thus, we regard the structured development of a taxonomy from a hindering point of view as important to identify, structure and overcome existing barriers to DT. Recently the research area of barriers is still dominated by unstructured lists which are hard to compare.

One of the first approaches to identify barriers is a study by Piatier from which the barriers to innovation approach has arisen [13]. Different classifications of barriers came up in literature since the approach was formulated. Often, these are differentiated in internal and external which are further subdivided. Internal barriers include resource-related, management-, time-, culture- and systems-related as well as human-related challenges. External barriers are subdivided into supply-related, demand-related and environmental-related [33]. The classification between internal and external has been proven as useful in many studies in different contexts [34, 35]. D'Este et al. [36] differ between revealed barriers describing the process of innovation and deterred barriers meaning the obstacles from adopting an innovation. In a study of Coad et al. from 2016 [37] items for four different barrier factors were used in a questionnaire: cost, knowledge, market, and regulation. Another differentiation has been pointed out in firm-related, project-related, product-related, and market-related factors [38]. This study also highlights the second side of the coin regarding barriers – success. Often success factors are described as the opposite from barriers. Success is more often in the center of research compared to the negative side of barriers [39]. Critical success factors are defined as "those few things that must go well to ensure success" [40]. The approaches are closely related to the adoption theory which deals with either adopting (success) or rejecting (failure) IT [41]. Although there is a long-lasting tradition of barrier research, a homogenous classification or guideline to describe barriers is still missing.

Especially in the dynamic field of digital transformation, classifications need to be developed. First studies identifying barriers in this context recently arise. The closely related "challenges" of DT are discussed in different approaches. These are, for example, strategic, organizational and cultural as well as implementation related [42]. Technical and market relevant challenges like time to market also exist [43]. Bilgeri and Wortmann [14] identified 16 barriers to innovation in an IoT context. They could highlight that DT also leads to new barriers referring to the knowledge of the people. The topic regarding business models is analyzed by different authors [14, 44]. Dremel [15] found six barriers to big data analytics for the automotive industry. In the context of the development of digital services through innovation contests, specific barriers like difficulties in finding competent team members are discussed [45]. So far, actual findings were valuable but they are generated in specific contexts. The results remain at a very detailed level and hardly allow systematic guidelines for action to be developed. They are heterogeneous and hard to compare, because they are often randomly compiled lists focusing on distinct aspects. Holistic approaches dealing with DT in a broader sense [46] are rather rare. We emphasize a certain need to characterize, describe and compare barriers to DT to be able to overcome these. A taxonomy is the first step to do so.

3 Method

As "fast-changing phenomena are difficult to investigate solely through the use of traditionally privileged methods” [47], it seems appropriate to conduct qualitative research in the context of DT. 46 interviews build the base for our study. In a first round, data from 30 participants were gathered. To check theoretical saturation [48], we gathered data from 16 more participants. Because no further impulses could be collected from the new data, we assessed the data set as useful for our purposes. We preceded the interviews in 31 different enterprises in varying industries. We aimed for a broad sample to be able to capture common base between different industries [49]. The taxonomy will be useful for a broad range of situations. Therefore, a purposeful sampling method was applied [50]. The selection criteria included a necessary affiliation of the interviewees to projects or departments explicitly dedicated to digital transformation in their company. All of the enterprises already implemented digital technologies successfully or/and were involved in digitalizing their processes and products. The dominant industries were automotive (7 companies) with mostly original equipment manufacturers (OEM) and agriculture (8 companies) with mostly agricultural machinery manufacturers (AMM). The length of the interviews ranges from 15 to 91 minutes with an average of 43 minutes. The interviewees are between 26 and 58 years old. Two interviewees are female and the rest male. All have German as mother tongue. Therefore, the interviews were conducted in German and mostly at the interviewee’s offices. Table 1 gives an overview of the participating interviewees. Industries, an identifier for case companies (to be able to associate the statements from interviews with the companies) and their role in the supply chain are shown. Many interviewees provided knowledge from the perspective of a user, producer and customer of digital products or services. The interviewees from the agricultural sector, for example, use digitalized production processes, sell machines with digital abilities and digital interfaces, and they buy digital services for big data applications which they use for their products and production processes.

All interviews were recorded, transcribed and translated for research purposes. An interview manual was used including (1) an introduction regarding digital transformation in general (including the understanding of the topic from the interviewee's perspective), (2) questions on the actual situation of DT and related barriers (including situations of success, frustration, emphasized support, and barriers), and (3) a subsumed report of three major barriers to DT.

Table 1. Overview of Interviewees

Industry	*Case*	*Role*
Automotive	Au1- Au7	OEM; supplier; service provider
Agriculture	AC1-AC8	AMM; food producer; farmer
Plastics Industry	P1-P3	processing chain; manufacturer
Steel Industry	SI1-SI3	processing chain
Other Manufacturing	OM1-OM4	engineering; cosmetics
Services	S1-S3	service provider; facility mgmt.; software solutions
Consulting	C1-C3	consultant

For the development of the taxonomy, we combined the approach according to Nickerson [18] with a qualitative content analysis [51]. We specified "meta-characteristics" by reading relevant literature regarding the field of barrier research. They serve as a guideline for the deduction of objects (Steps 4e – 6e) from the empirical material gained from the interviews. We used empirical data as the main source to develop the taxonomy. However, results from the literature were used to complement the understanding. The steps from Nickerson as well as our application for this study is shown in Table 2.

As an approach, we mainly used the empirical-to-conceptual procedure to gain insights on the characteristics (C) first. Afterward, we clustered the obtained characteristics according to dimensions (D). We openly coded the interviews [52] to iteratively develop a well-prepared database. We used the online-tool QCAmap for coding the material. Statements about barriers to DT and similarities regarding the structure and possible characteristics of the barriers were searched for. We sequentially adjusted the research for a simultaneous collection of data and analysis. All authors were permanently involved in the research process of data generation and analysis. Thus, making it necessary to discuss the coding in which the coders disagreed to come to an inter-coder agreement [51]. This led to a repetitive induction and discussion of concepts [53]. The cross-fertilized relationship between analysis and data generation as well as the development of the taxonomy led to a complex process of iterative revisions [54]. The research process was characterized by a comparative method [52] to deduce logical and consistent categories of barriers to digital transformation. Besides the data from the interviews, we included corresponding research streams for the development of the taxonomy. These also implied a conceptual-to-empirical iteration which complements the procedure.

Table 2. Development of the Taxonomy

Step by Nickerson	*Application of Steps*
Step 1: Determine meta-characteristic	(1) who is affected by the barrier; (2) major implications; (3) next steps for barrier
Step 2: Determine ending conditions	Going in iterations; no new, merged or split D or C in the last iteration; at least two Cs under every D
Step 3: Choose approach	Mostly empirical-to-conceptual
Step 4e: Identify (new) subset of objects	Coding of barriers and/or identification of barriers from literature
Step 5e: Identify common characteristics and group objects	Check context of barriers in interviews and/or look for characteristics in literature
Step 6e: Group characteristics into dimensions to create (revise) characteristics	Name dimensions independently and discuss within researcher group
Step 7: Ending conditions met?	Check independently by all authors; only if all agree, the process ends

Nickerson does not intend the last step of our research approach. Finally, we evaluated the taxonomy on the base of a qualitative interview. We chose a sample of ten interviewees and asked the interview-partners to apply the taxonomy with the help of a self-chosen case example. Afterward, we asked for an assessment and explored additional dimensions. The sample consisted of practical users (mainly IT consultants) and researchers in the field of DT. None of the participants was involved in the constructive analysis of the model or the creation of the taxonomy. As the sample of ten is rather small, we regard the evaluation as ongoing.

4 Taxonomy of Digital Transformation Barriers

During our research procedure, we deduced nine different dimensions (Table 3) that help to describe and explain the barriers to DT. The dimensions provide none or just parsimonious overlapping.Some dimensions are slightly connected (e.g., effect and duration). Characteristics describe dimensions. They explain the quality of the barriers. Some dimensions seem obvious to describe barriers to DT (e.g., stakeholder, impact and IT integration). Others needed a careful analysis and interpretation of the context (such as visibility, nature, solvability, duration, effect, recipient). Any dimension alone can be used to cluster or group existing barriers. However, the interplay of dimensions delivers a multi-focal view of the problem. The taxonomy helps to find out who is actually affected by the barrier (stakeholder, impact), what are the major implications (IT-integration, visibility, nature, solvability, integration) and what is the possible future of the problem (duration, recipient). Furthermore, we try to align the dimensions to related research streams.

Table 3. Taxonomy of DT Barriers

Dimension	*Characteristics*		
Stakeholder Who emphasizes the barrier?	individual	organiza- tional	environ- mental
Impact Which level does the barrier impact?	internal		external
IT-Influence How strong is the influence of IT?	social	socio- technical	technical
Visibility How obvious is the barrier?	hidden		visible
Nature What does the stakeholder sense?	standing by	fear of	lack of
Solvability What is the actual status of the problem?	low		high
Effect How huge is the effect on development?	no	slow	stop
Duration How long will the barrier last?	short-term	long-term	permanent
Recipient Who may solve the problem?	inner circle		outer circle

Stakeholder. The dimension stakeholder shows who emphasizes the barrier. The differentiation between the stakeholder groups is inspired by Coad et al. [37] who concentrated more on the level of who/what is affected. The stakeholder is not necessarily the person who expresses the barrier. Barriers often result from a lack of adoption. Adoption proceeds on different levels [55]. The level of impact is therefore related to the individual [56], the organization [57, 58] or the surrounding environment [59]. “Especially the internal employees. They are left alone. They do not know how they can handle this on their own.” (AC8) The characteristic organizational implies the core-enterprise as well as corporate networks for development and knowledge transfer. “It is all about cooperation. If data are not shared, the [digital transformation] won’t be a success.” (Au2) Customers [43] and suppliers (“The more change, the more risk to fail. That is, what the customer realizes” (Au3)), but also governmental or other third party organizations are subsumed with the term environmental.

Impact. Barriers to digital transformation do not only affect the stakeholders themselves but their surroundings as well [13, 34]. It is a question of where the barrier occurs. Therefore, we identified two contrasting forms of impact. The internal view, including employees, departments or the whole company, can hinder the pervasion. Internal change resistance [60], fear of high investments [36, 45] or lack of skills [61] proceed here. "We have a change in skills needed.” (Au5) However, digital transformation also implies the blur of borders and forces cooperation on a business level. “Network integration will be the biggest problem.” (AC1) Therefore, often external partners, customers, suppliers or others are also affected [8]. “Soon the digital transformation will change the market from buying things to using data. This will affect the suppliers.” (C2) Still, enterprises complain about the non-existence of legal standards. Using their market power, companies can develop quasi-standards in the new markets of DT. “Currently, in the software development, enterprises define the interfaces. […] I expect problems of norming occurring in the future." (OM2) Due to blurring borders, “internal” might mean more than just one company, e.g. in the evolvement of innovation networks [62].

IT-Influence. Digital transformation is to some extent a process of IT adoption [15] to achieve success [63]. Hence, barriers to DT can affect different stages of human-computer interaction caught between technical, socio-technical and social barriers. The pure technical barriers often result from a lack of technical standards [64], security aspects [65], and integration to the existing system environment [66]. “That is the problem. Digital transformation means: to work in a big general store with hundreds of different technologies and solutions.” (Au2) Furthermore, the access to network technologies slows down the DT. “One problem is the lack of network speed at specific locations.” (Ac1) The socio-technical implications refer to the knowledge of the workers and the way the technology is used [67]. “They left the people on the shop floor alone. That reduces the quality. There was no one left, who cared for the problems because they were all in front of their computers.” Moreover, the clear social implications are triggered by the DT but relate more to social effects like the fear of unemployment [2]. “Yes, I guess, many workers are afraid, that digitalization of works implies the loss of the workplace.” (AC8)

Visibility. This dimension describes how obvious a barrier is for third parties. Reasons for the hiding of barriers differ. The fear of repercussions, avoiding conflicts and even anxiety of job loss and harassment play an important role [68]. Especially barriers that rely on personal problems and estimations are slightly invisible. “Often the people do not tell. However, they are afraid.” (S2) Talking about barriers increases visibility. Furthermore, technical problems can be hidden and only appear when they accumulate. However, visibility is dependent on the observer. What is obvious for one group may be hidden from the other. Visibility, as well as existing barriers, can be very dynamic. "You cannot predict when the feeling of employees regarding the project changes, but suddenly it does.” (C2) Visible and invisible barriers have also been observed in case studies in research [69].

Nature. To understand what the stakeholders perceive and what they sense eases the later solvation. We differentiate three manifestations: standing by, fear of and a perceived lack of. The transition between these characteristics is often seamless. At an early stage of change for innovation, people and organizations prevent from changing existing structures. This behavior can be named resistance to change or the desire to stand by and can also be found in innovation research [70, 71]. “Possessive thinking, traditional role administration, […] they hinder the process.” (Au5) Many barriers also result from deficiencies. “The lack of standards is one of the major barriers to network integration. Moreover, of course, the habit of interdisciplinary co-working.” (Au5) For many enterprises, the implied investment and the lack of free capital plays an important role. “Initially it is expensive, and we do not save money at first glance.“ (Au1) Moreover, we detect a large group of unspecific fears that hinder the successful digital transformation. “There is much diffuse fear. Diffuse is important in this context.” (Au5)

Solvability. Solvability describes the actual status of the problem or barrier. Do the subjects of the obstacle consider the impediments as solvable or not from the current status? The classification of a barrier within this dimension might change over time as the status of its solvability might change. There are for example barriers that are (from the actual time of consideration) not solvable, such as standardization of interfaces and legal policies [72]. “From my point of view, we suffer from deficits of the technology. To name it: we have striking lack of standards.” (C1) Then again, other problems are more easily to solve, such as missing skills. “We need to […] train future skills.“ (SI1)

Effect. It is important whether the barrier has the potential to lead to massive interferences of an enterprises’ DT. We gradually distinguish the effects regarding how massive the barrier affects the DT process. Some implications seem to be negligible while others may stop the digital transformation process. A missing legal base for the use of data and the interconnection of firms can harm the DT of a company’s network [73]. “There is no improvement in sight. We realize that some things like works council, data protection statements massively brake the development.” (S2) Otherwise, barriers like a lack of technology acceptance from a few employees are threatening but not existential. “We worked with the employees. We went there and took a look at the situation. Why didn’t it work? Moreover, we included them in the development process.” (P3) There are also several research contributions dealing with the nature of technology acceptance [56, 71] and ways to handle [74].

Duration. The estimation of the duration of the barrier is also necessary to consider for the future planning of DT. The length of the barrier may influence strategic planning [10]. “What are we doing? We have to change as an organization. Moreover, this is difficult: to change the whole business model.“ (AC2) Many interviewees see the DT mainly as a long-term challenge: “There is still a lot to do to find the right solutions.” (Au4) Research about DT maturity may help to assess the duration of change processes [72]. We chose the dimension duration rather than a dimension that aims at the processual aspect including different states of adoption (suggested by Henriette et al. [42]). We consider the differentiation of adoption stages as less applicable, as the stages are often overlapping and they differ even within the enterprises as rollouts often proceed in several steps.

Recipient. Considering there is a possible solution for the barrier, it is central to understand who may solve the problem. We differentiate the inner and outer circle [75]. While the inner circle reflects the changing organization, its parts, and partners, the outer circle comprises legal institutions and governments. It is often up to the people to foster the DT. “I like to join community meetings. […] We share knowledge, and we exchange views.” (AC2) We assume that there will often be a group of people and institutions who remove barriers [42]. The overcoming of barriers is seldom limited to individuals. Schools (“It is a major barrier: the skills. IT should start in school.” (AC8)), governments (“The biggest barrier is the whole legal thing.“ (Au2)), and the people have to work together. „And it is due to the people, the organizations, and institutions beyond our organization like trade associations to become friends with the whole topic.” (Au5)

5 Discussion and Future Research

The evaluation of the model confirms the usability of the taxonomy. None of the respondents had a problem naming a barrier for DT and describing the problem using dimensions and characteristics. Using a 4-point-Likert-scale (very useful, useful, not useful, annoying), most respondents rated the model as useful or very useful for identifying, describing and discussing the barriers. However, the taxonomy is considered only slightly useful to find solutions. The group of consultants requested more space for solutions (e.g., for the resources needed to solve the problems) and the way to communicate the barrier. The group of researchers suggested further gradations for the characteristics. All respondents confirmed the high usability for the discussion of barriers. It will be part of future research to find out how the taxonomy can be used to identify barriers within DT projects. The taxonomy can be enhanced with dimension regarding the salvation of problem. Since the evaluation is not yet complete, we aim for future research at questions like: Which industry-specific adjustments of the taxonomy have to be made? How can the identification of barriers promote the formulation of a DT strategy [10]?

By using this taxonomy, one can develop a profile of any barrier which can help to overcome these challenges. For example, a wide discussed barrier to DT is the fear of losing jobs [2] which is often emphasized internally by individuals. It is a social barrier which is in many cases not visible to others. This could also mean that the solvability is complicated. It could affect others and last permanently. The inner or outer circle could be responsible for salvation. These interpretations of the barrier are not fixed. They depend on the company and the specific barrier that hampers DT. So, the taxonomy is useful in practice, if it is applied by companies to identify characteristics of barriers and thus be able to find solutions. In this example, employees would need to be educated about further projects and their impacts.

Though we carefully conducted this research, there are still limitations to our work. Even though the holistic view implies advantages regarding the usefulness of the taxonomy, it is also a limitation because the object of analysis is broad. Moreover, the qualitative approach implies limitations by nature as results could be due to the sample and subjectivity of the researchers. We tried to minimize this by coding interview data independently and discussing results in iterative rounds. Furthermore, the sample was chosen broadly to cover many different opinions. Besides this qualitative approach, it could be very fertile to quantify the characteristics of the dimensions. Moreover, it could be of great value to combine the findings of this study with case studies focusing on the overcoming of barriers. The evaluation and quantitative verification of the taxonomy is a call for additional research in this area.

6 Conclusion and Contribution

Any barrier in the field of DT leads to a slow down or complete termination of the digital change in enterprises. It is of great importance for research and practice to understand barriers to digital transformation. Different lists of barriers already exist in research but it is hard to comprehend those. To understand the barriers from a broader perspective, we developed a taxonomy that implies dimensions and characteristics of these to describe and classify barriers. The taxonomy consists of nine dimensions. Each of them answers questions on barriers to DT. Potential barriers can be classified by using each dimension of the taxonomy. Then, the characteristics can be used to describe barriers more in detail. The taxonomy was the result of a research process that orients towards Nickerson`s approach [18]. The findings of this study rely on the base of a qualitative analysis of 46 semi-structured interviews. To ensure the quality and validity of our conclusions, we enriched the results by further research on barriers.

The first and still ongoing evaluation of model proves the usability of the model so far. The structured taxonomy helps to identify, describe and discuss barriers to DT in research and practice. The taxonomy can foster the creation of a knowledge base for the ongoing process of DT [4]. Practitioners can use the taxonomy to understand and detect obstacles of DT as well as deduce action plans for a successful transformation. If firms can reflect upon their challenges, countermeasures can be taken.

Our approach builds upon the barriers to innovation literature by integrating barrier types like internal and external barriers [33] but adding new dimensions like visibility and solvability to structure barriers. Researchers in the field of barriers to DT research [e.g., 14, 15] can use the taxonomy to classify the identified barriers. Future findings can easily be aligned with actual and further research and connections can be discovered. We hope to foster the discussion on the understanding and structuring barriers.

References

1. Schweer, D., Sahl, J.C.: The Digital Transformation of Industry – The Benefit for Germany. In: Abolhassan, A. (ed.) The Drivers of Digital Transformation, pp. 23–31. Springer, Cham (2017)
2. Brynjolfsson, E., McAfee, A.: Race Against the Machine. Digit. Front., Lexington (2011)
3. Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M.: Embracing Digital Technology: A New Strategic Imperative. MIT Sloan Manag. Rev. 55, 1–12 (2014)
4. Schwab, K.: The Fourth Industrial Revolution. Penguin UK, London (2017)
5. Morakanyane, R., Grace, A., O'Reilly, P.: Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature. In: Proceedings of Bled eConference, pp. 427–443, Bled (2017)
6. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for Implementing the Strategic Initiative Industrie 4.0. acatech – National Academy of Science and Engineering (2013)
7. Matt, C., Hess, T., Benlian, A.: Digital Transformation Strategies. Bus. Inf. Syst. Eng. 57, 339–343 (2015)

8. Banker, R., Mitra, S., Sambamurthy, V.: The Effects of Digital Trading Platforms on Commodity Prices in Agricultural Supply Chains. Manag. Inf. Syst. Q. 35, 599–611 (2011)
9. Benlian, A., Hilkert, D., Hess, T.: How Open is this Platform? The Meaning and Measurement of Platform Openness from the Complementors' Perspective. J. Inf. Technol. 30, 209–228 (2015)
10. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for Formulating a Digital Transformation Strategy. MIS Q. Exec. 15, 123–139 (2016)
11. Lerch, C., Jäger, A., Maloca, S.: Wie digital ist Deutschlands Industrie wirklich? Arbeit und Produktivität in der digitalen Produktion. Arbeitsbericht, Fraunhofer ISI, Ausgabe 71, Karlsruhe (2017)
12. Hirsch-Kreinsen, H.: Digitization of Industrial Work: Development Paths and Prospects. J. Labour Mark. Res. 49, 1–14 (2016)
13. Piatier, A.: Barriers to Innovation. Pinter Pub Ltd., London (1984)
14. Bilgeri, D., Wortmann, F.: Barriers to IoT Business Model Innovation. In: Proceedings of Wirtschaftsinformatik, St.Gallen (2017)
15. Dremel, C.: Barriers to the Adoption of Big Data Analytics in the Automotive Sector. In: Proceedings of AMCIS, Boston (2017)
16. Glass, R.L., Vessey, I.: Contemporary Application-Domain Taxonomies. IEEE Softw. 12, 63–76 (1995)
17. Weking, J., Stöcker, M., Kowalkiewicz, M., Böhm, M., Krcmar, H.: Archetypes for Industry 4.0 Business Model Innovations. In: Proceedings of AMCIS, New Orleans (2018)
18. Nickerson, R.C., Varshney, U., Muntermann, J.: A Method for Taxonomy Development and its Application in Information Systems. Eur. J. Inf. Syst. 22, 336–359 (2013)
19. Bowen, G.A.: Grounded Theory and Sensitizing Concepts. Int. J. Qual. Methods. 5, 12–23 (2006)
20. Erol, S., Schumacher, A., Sihn, W.: Strategic Guidance towards Industry 4.0 – A three-stage Process Model. In: Proceedings of the International Conference on Competitive Manufacturing, pp. 495–502. Stellenbosch (2016)
21. Püschel, L., Röglinger, M., Schlott, H.: What's in a Smart Thing? Development of a Multi-Layer Taxonomy. In: Proceedings of ICIS, Dublin (2016)
22. Loonam, J., Eaves, S., Kumar, V., Parry, G.: Towards Digital Transformation: Lessons Learned from Traditional Organizations. Strateg. Change. 27, 101–109 (2018)
23. Sanchez, M.A.: A Framework to Assess Organizational Readiness for the Digital Transformation. Dimens. Empres. 15, 27–40 (2017)
24. Rizk, A., Bergvall–Kåreborn, B., Elragal, A.: Towards a Taxonomy for Data-Driven Digital Services. In: Proceedings of HICSS (2018)
25. Akram, A., Åkesson, M.: A Research Framework to Study How Digital Service Innovation Transforms Value Networks. In: Proceedings of Information Systems Research Seminar in Scandinavia (IRIS), Turku, pp. 29–40 (2011)
26. Chowdhury, S., Akesson, M.: A Proposed Conceptual Framework for Identifying the Logic of Digital Services. In: Proceedings of PACIS, Brisbane (2011)
27. Bondar, S., Hsu, J.C., Pfouga, A., Stjepandić, J.: Agile Digital Transformation of System-of-Systems Architecture Models Using Zachman Framework. J. Ind. Inf. Integr. 7, 33–43 (2017)
28. Boban, M., Weber, M.: Internet of Things, Legal and Regulatory Framework in Digital Transformation from Smart to Intelligent Cities. In: International Convention on Information and Communication Technology, Electronics and Microelectronics (2018)

29. Kutzner, K., Schoormann, T., Knackstedt, R.: Digital Transformation in Information Systems Research: A Taxonomy-Based Approach to Structure the Field. In: Proceedings of ECIS, Portsmouth (2018)
30. Heilig, L., Lalla-Ruiz, E., Voß, S.: Digital Transformation in Maritime Ports: Analysis and a Game Theoretic Framework. NETNOMICS: Econ. Res. Electron. Netw. 18, 227–254 (2017)
31. Li, F.: The Digital Transformation of Business Models in the Creative Industries: A Holistic Framework and Emerging Trends. Technovation, in press (2018)
32. Liere-Netheler, K., Vogelsang, K., Packmohr, S., Hoppe, U.: Towards a Framework for Digital Transformation Success in Manufacturing. In: Proceedings of ECIS, Portsmouth (2018)
33. Hadjimanolis, A.: The Barriers Approach to Innovation. In: Shavinina, L.V. (ed.) The International Handbook on Innovation. pp. 559–573. Pergamon, Oxford (2003)
34. Demirbas, D., Hussain, J.G., Matlay, H.: Owner-Managers' Perceptions of Barriers to Innovation: Empirical Evidence from Turkish SMEs. J. Small Bus. Enterp. Dev. 18, 764–780 (2011)
35. Madrid-Guijarro, A., Garcia, D., Van Auken, H.: Barriers to Innovation Among Spanish Manufacturing SMEs. J. Small Bus. Manag. Milwaukee. 47, 465–488 (2009)
36. D'Este, P., Iammarino, S., Savona, M., von Tunzelmann, N.: What Hampers Innovation? Revealed Barriers Versus Deterring Barriers. Res. Policy. 41, 482–488 (2012)
37. Coad, A., Pellegrino, G., Savona, M.: Barriers to Innovation and Firm Productivity. Econ. Innov. New Technol. 25, 321–334 (2016)
38. van der Panne, G., van Beers, C., Kleinknecht, A.: Success and Failure of Innovation: A Literature Review. Int. J. Innov. Manag. 7, 309–338 (2003)
39. Distel, B., Ogonek, N.: To Adopt or not to Adopt: A Literature Review on Barriers to Citizens' Adoption of E-Government Services. In: Proceedings of ECIS, Istanbul (2016)
40. Boyton, A.C., Zmud, R.W.: An Assessment of Critical Success Factors. Sloan Manage. Rev. 25, 17–27 (1984)
41. Rogers, E.M.: Diffusion of Innovations. Free Press, New York. Collier Macmillan, London (1983)
42. Henriette, E., Feki, M., Boughzala, I.: Digital Transformation Challenges. In: Proceedings of the Mediterranean Conference on Information Systems (MCIS), Paphos (2016)
43. Gebhart, M., Giessler, P., Abeck, S.: Challenges of the Digital Transformation in Software Engineering. In: Proceedings of the International Conference on Software Engineering Advances (ICSEA) (2016)
44. Chesbrough, H.: Business Model Innovation: Opportunities and Barriers. Long Range Plann. 43, 354–363 (2010)
45. Hjalmarsson, A., Johannesson, P., Jüll-Skielse, G., Rudmark, D.: Beyond Innovation Contests: A Framework of Barriers to Open Innovation of Digital Services. In: Proceedings of ECIS, Tel Aviv (2014)
46. Karimi, J., Walter, Z.: The Role of Dynamic Capabilities in Responding to Digital Disruption: A Factor-Based Study of the Newspaper Industry. J. Manag. Inf. Syst. 32, 39–81 (2015)
47. Sarker, S., Xiao, X., Beaulieu, T.: Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles. MIS Q. 37, iii–xviii (2013)
48. Strübing, J.: Was ist Grounded Theory? In: Strübing, J. (ed.) Grounded Theory. pp. 9–35. Springer, Wiesbaden (2014)
49. Yin, R.K.: Case Study Research: Design and Methods. SAGE, Los Angeles (2014)
50. Marshall, M.N.: Sampling for Qualitative Research. Fam. Pract. 13, 522–525 (1996)

51. Mayring, P.: Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. SSOAR, Klagenfurt (2014)
52. Glaser, B.G., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Publ. Co., Chicago (1975)
53. Corbin, J., Strauss, A.: Grounded Theory Research: Procedures, Canons and Evaluative Criteria. Z. Für Soziol. 19, 418–427 (1990)
54. Strauss, A., Corbin, J.: Basics of Qualitative Research. Newbury Park, CA, Sage (1990)
55. Zhu, K., Dong, S., Xu, S.X., Kraemer, K.L.: Innovation Diffusion in Global Contexts: Determinants of Post-Adoption Digital Transformation of European Companies. Eur. J. Inf. Syst. 15, 601–616 (2006)
56. Davis, F.D.: A Technology Acceptance Model for Empirically Testing New End-User Information Systems. Thesis (Ph.D.), Massachusetts Institute of Technology, Sloan School of Management (1986)
57. Gable, G.G., Sedera, D., Chan, T.: Re-Conceptualizing Information System Success : The IS-Impact Measurement Model. J. Assoc. Inf. Syst. 9, 377–408 (2008)
58. DeLone, W.H., McLean, E.R.: Information Systems Success Measurement. Found. Trends® Inf. Syst. 2, 1–116 (2016)
59. Tornatzky, L.G., Fleischer, M.: The Processes of Technological Innovation. Lexington Books, Lexington (1990)
60. Rafferty, A.E., Jimmieson, N.L.: Subjective Perceptions of Organizational Change and Employee Resistance to Change: Direct and Mediated Relationships with Employee Well-being. Br. J. Manag. 28, 248–264 (2017)
61. Pinzone, M., Fantini, P., Perini, S., Garavaglia, S., Taisch, M., Miragliotta, G.: Jobs and Skills in Industry 4.0: An Exploratory Research. In: Lödding, H., Riedel, R., Thoben, K.-D., von Cieminski, G., Kiritsis, D. (eds.) Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing. pp. 282–288. Springer, Cham (2017)
62. Galvagno, M., Dalli, D.: Theory of Value Co-Creation: A Systematic Literature Review. Manag. Serv. Qual. Int. J. 24, 643–683 (2014)
63. Holotiuk, F., Beimborn, D.: Critical Success Factors of Digital Business Strategy. In: Proceedings of Wirtschaftsinformatik. pp. 991–1005, St. Gallen (2017)
64. Pfeiffer, T., Hellmers, J., Schön, E.M., Thomaschewski, J.: Empowering User Interfaces for Industrie 4.0. Proc. IEEE. 104, 986–996 (2016)
65. Moody, G.D., Siponen, M., Pahnila, S.: Toward a Unified Model of Information Security Police Complience. MIS Q. 42, 285-A22 (2018)
66. Melville, N., Kraemer, K., Gurbaxani, V.: Review: Information Technology and Organizational Performance: An Integrative Model of It Business Value. Manag. Inf. Syst. Q. 28, 283–322 (2004)
67. Brown, M.M.: Technology Diffusion and the "Knowledge Barrier." Public Perform. Manag. Rev. 26, 345–359 (2003)
68. Ryan, K., Oestreich, D.K.: Driving Fear out of the Workplace: How to Overcome the Invisible Barriers to Quality, Productivity, and Innovation. Jossey-Bass, San Francisco (1991)
69. Lee, S.M.: A Comparative Study of the Automobile Industry in Japan and Korea: From Visible to Invisible Barriers. Asian Surv. 51, 876–898 (2011)
70. Kim, H.-W., Kankanhalli, A.: Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. MIS Q. 33, 567-582 (2009)

71. Polites, G.L., Karahanna, E.: Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. MIS Q. 36, 21–42 (2012)
72. Berghaus, S., Back, A.: Stages in Digital Business Transformation: Results of an Empirical Maturity Study. In: Proceedings of the Mediterranean Conference on Information Systems (MCIS), Paphos (2016)
73. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. Future Gener. Comput. Syst. 29, 1645–1660 (2013)
74. Vogelsang, K., Steinhüser, M., Hoppe, U.: A Qualitative Approach to Examine Technology Acceptance. In: Proceedings of ICIS, Milan (2013)
75. Sayre, S.: Qualitative Methods for Marketplace Research. Sage, Thousand Oaks (2011)

Ambidexterity in Service Innovation Research: A Systematic Literature Review

Verena Wolf

Paderborn University, Department Wirtschaftsinformatik, Paderborn, Germany
verena.wolf@upb.de

Abstract. Increased interconnectedness of multiple actors and digital resources in service eco-systems offer new opportunities for service innovation. In digitally transforming eco-systems, organizations need to explore and exploit innovation simultaneously, which is defined as ambidexterity. However, research on ambidextrous service innovation is scarce. We provide a systematic literature review based on the concepts of ambidexterity, offering two contributions. First, research strands are disconnected, emphasizing either exploration or exploitation of service innovation, despite an organizations' need to accelerate innovation cycles of exploring and exploiting services. Second, a new framework for ambidextrous service innovation is provided, inspired by the dynamism and generative mechanisms of the ontologically related concept of organizational routines. The framework adopts the perspective of a mutually constitutive relationship between exploring new and exploiting current resources, activities, and knowledge. The findings remedy the scattered literature through a coherent perspective on service innovation that responds to organizations' needs and guides future research.

Keywords: Exploration, Exploitation, Service Innovation, Organizational Routines, Ambidexterity

1 Introduction

Digital transformation is a buzzword [1], which describes "a technology-induced change […] that includes both the exploitation of digital technologies to improve existing processes, and the exploration of digital innovation" [2]. Digital technologies, i.e. devices that process a binary computational code and smart technologies, i.e. interconnected devices that share information and interact with its users and other devices [3], provide opportunities for innovative services [4]. A service is designed through a collaborative process by integrating knowledge and skills (operant) and tangible resources (operand) that provide value to actors in a service system [5]. A fundamentally new process or service offering—either in addition to current services or as change in the delivery process—is defined as a service innovation (SI).

The complexity and rapid speed of changing markets and technologies [6] increase competition as well. Organizations need to accelerate cycle-times of exploring new SI and exploiting them efficiently [7]. The management of dual capacities of exploring

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

and exploiting is defined as ambidexterity [8]. Exploration is associated with radical SI to attain progress in changing service eco-systems [8]. However, new edgy services may not always be appealing to customers or take time to unleash their full potential, which can be risky for organizations [9]. While exploring SI, organizations need to exploit existing services through continuous improvement and increasing the efficiency of service processes [8]. Organizations like MySpace, studyVZ, or Vine that only focused on exploitation without exploring new value-adding features for their service portfolio have experienced declining user numbers and finally had to shut down their businesses. These cases demonstrate that organizational ambidexterity is an important capability in dynamic and digitally transforming service eco-systems [10].

Whilst organizations need to focus on exploring and exploiting SI simultaneously, research on SI is getting increasingly extensive and dispersed [11]. Previous literature reviews have pointed out that SI encompasses different aspects, such as New Service Development, Service (Systems) Engineering, and Service Management [6]. However, an integrated view on ambidextrous SI seems to be scarce [12], motivating our research question: "*To what extent are exploration and exploitation of SI covered in research?*"

The paper offers two main contributions. First, we identify a disjunction of SI literature with articles either focusing on exploration or exploitation. Second, we provide an integrative conception of the different research strands that uncovers the fundamental mechanisms of SI. We derive three propositions from a conceptual analysis and develop a framework for ambidextrous SI. The framework builds on dynamism and generative mechanisms, inspired by the concept of organizational routines—hereafter referred to as routines only. Routines are "recognizable patterns of interdependent actions, carried out by multiple actors" [13]. SI and routines are ontologically related since both are performed by multiple actors, are composed of some recognizable patterns of activity, and are endogenously changing. Further, the paradox of simultaneous competing pressures for exploration and exploitation activities has already been resolved in routines literature [6], helping to understand ambidexterity in SI as well and thereby, remedying the disjunction in SI literature.

The paper unfolds as follows: In Section 2, theoretical foundations of SI and ambidexterity are provided. Section 3 justifies the systematic literature review as research method. In Section 4 the results from the literature review are described and presented in a concept matrix. Section 5 continues with the discussion of insights resulting from the concept matrix. We derive three propositions that are used to develop a framework, which unifies scattered SI research through ambidexterity. The paper is concluded in Section 6 with the contributions to research, implications, and limitations.

2 Theoretical Foundations

2.1 Service Innovation Research

Service science is the study of service systems [14], which focuses on the co-creation of value within complex constellations of integrated resources, leading to innovative services [14, 15]. A service system is the basic unit of analysis [14], which describes "a configuration of people, technologies, and other resources that interact with other service systems to create mutual value'' [16]. Value emerges through the use and application of resources within a service system [5]. SI emerged as a new concept during the 1980's in service science research [17]. At the same time, New Service Development research emerged in the Anglo-American literature. Soon thereafter, Service Engineering arose in Germany as another research stream that adopted approaches from product engineering to design a value proposition in a structured way [18]. Whereas Service Engineering encompasses models, methods, and principles to design individual services, Service Systems Engineering describes the design and development of another basic unit of analysis, i.e. service systems as integrated conglomerates of services, products, and information technology [16, 19]. SI, New Service Development, and Service (Systems) Engineering emanate from different research disciplines, but are often used interchangeably [20].

There are two alternative views for conceptualizing value and value creation, which are referred to as "value-in-exchange" and "value-in-use" [15]. Value-in-exchange refers to the output or distribution of service, focusing on a goods-dominant logic (GDL) view [15]. Value-in-use is defined in service-dominant logic (SDL) by describing that value is determined by a user's consumption [21]. Value is co-created jointly and reciprocally by multiple actors through the integration of resources and application of competences [15]. Hence, value evolves differently in exploring and exploiting products and SI. In the light of SDL, we refer to service as "the application of specialized competences (knowledge and skills) through deeds, processes, and performances for the benefit of another entity or the entity itself' [5].

In a digitally transforming eco-system, organizations need to engage in accelerating cycles of SI to expedite growth, improve quality of services, satisfy customer needs, and to be competitive [17]. SI describes both, a change of an existing value-proposition or a design of a new value-proposition [22]. This includes the configuration or reconfiguration of resources within a service system [23]. Innovation does not only create value for an organization, it also changes the eco-system in which an organization operates [9]. Hence, SI needs to be distinguished from an invention, since an invention has no inherent value [9].

The process of SI can be planned, as an outcome of development activities or emerge from an individual's actions [24]. A multitude of methods for planning and developing innovative services have been proposed [18]. Whereas initial methods suggested sequential or linear aligned steps to develop innovative services, more recent models rather propose iterative patterns for service design [17, 18]. Other research focuses on managerial frameworks with a focus on different resources, information, and knowledge for SI [4, 12, 17, 22, 23].

2.2 Ambidextrous Organizations in Digital Transformation

Organizations need to adapt to a changing eco-system and technological change [25], which causes conflicting requirements. On the one hand, organizations have to achieve innovation through the exploration of new competencies [26]. On the other hand, they have to be efficient by exploiting their current capabilities [26]. In early research, the activities of exploration and exploitation have been seen as an incompatible trade-off [27]. However, focusing only on exploration will lead to low returns, whereas organizations that engage only in exploitation will eventually become deprecated [28].

More recent research introduced the concept of ambidextrous organizations. Ambidexterity refers to an organization's ability to be adaptive to changes for long-term success (*exploration*) whilst reducing variance and leveraging existing resources and capabilities efficiently (*exploitation*) [28]. Exploration and exploitation need to be well-balanced for the current continuance and future viability of an organization [25].

Until the late 1990's, ambidexterity was used to describe dual structures in organizations [29]. Some units of the organization were focused on searching and (re-) combining resources to achieve variation, while other units were employed for efficient alignment of organizational structures [26]. Ambidexterity was conceptualized as a temporal sequencing of exploitation and exploration [8]. However, in fast changing eco-systems, a periodically change by adapting new strategies and structures is ineffective [30]. Organizations need to engage simultaneously in exploitation ensuring current viability and devote sufficient attention to exploration to ensure the future viability [31]. In recent research, contextual ambidexterity has become prevalent, which is manifested in specific actions of individuals in organizations [29]. Contextual ambidexterity is the capacity to simultaneously demonstrate exploration and exploitation in each entity of an organization [29]. Since digital transformation requires removing organizational silos that work autonomously from other units [32], contextual ambidexterity is a crucial capacity of organizations. Thus, organizations continuously streamline current business activities and engage in innovation activities [8].

3 Research Method

The systematic literature review on SI is based on the methodological guidelines by Webster and Watson [33] and vom Brocke et al. [34]. We refer to Cooper's [35] taxonomy to cover a sufficient and representative degree of SI research (*Figure 1*) [34]. The focus is on both, methods and theories of SI since they are mutually constitutive. The goal is to gain a better understanding of the central concepts and thereby, identify issues and suggesting areas for further progress in the SI literature [35].

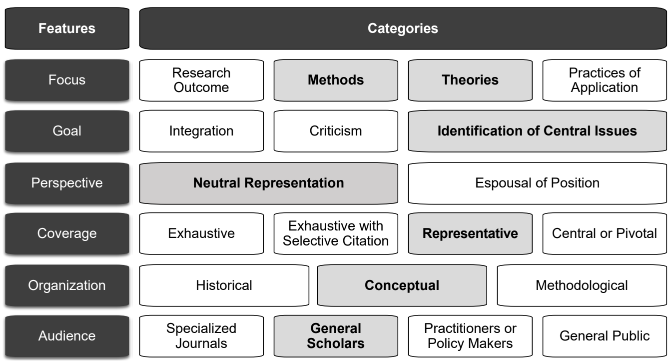


Figure 1. Positioning the Literature Review [35]

The perspective adopted for this literature review is neutral to provide a realistic representation of the current literature. A representative coverage of the literature by discussing the characteristics of SI makes the sample illustrative for a larger group of literature [35]. Although the audience of the review focuses mainly on IS scholars, we address researchers in service science, too. We conducted an iterative process for our literature search by defining and re-defining a search string and query structures through the course of the study. We started with scanning the literature to identify key terminology and synonyms. Since SI, New Service Development, and Service (Systems) Engineering emanate from different research disciplines, whose terms are often used interchangeably [34], we added the terms to our search string (see *Table 1*).

Table 1. Search String

("service" OR "service system" OR "services" OR "service science") AND ("innovation" OR "innovate" OR "innovativeness" OR "innovating" OR "invention" OR "creation" OR "create" OR "engineering" OR "design" OR "designing" "development" OR "develop" OR "new" OR "novel" OR "configuration" OR "configure" OR "organization" OR "organizing" OR "management" OR "process" OR "technique" OR "value cocreation" OR "value co-creation" OR "value proposition" OR "outcome" OR "resource" OR "exploration" OR "explore" OR "exploitation" OR "exploit")

Major contributions are published in leading journals and some conference proceedings [33]. Hence, we started by scanning the table of content and performed key word searches within these publications. The journals and conference proceedings were selected interdisciplinary, originating from the IS and Management discipline and were ranked as A+, A, or B according to the "VHB-JOURQUAL 3" ranking. Then, we extended the search within literature databases of AISeL, ProQuest, EBSCOhost, Google Scholar, Science Direct, and Web of Science. Afterwards, we conducted a backwards and forwards search to identify additional articles [33]. In total, the search yielded 1,145 articles. Duplicates were removed and articles that did not adhere to the following criteria were excluded: written in English, published in peer-reviewed scholarly journals or conference proceedings, address the focus mentioned in the taxonomy[1], and are available in full-text. Then, we performed an analysis of the title, keywords, abstract, and browsed the papers to identify, which articles are relevant for our review, leading to a sub-sample of 51. Lastly, we excluded research that was ranked lower than B level, narrowing down the literature to a final sample of 25 articles.

We followed the sequential model for structured content analysis by Mayring [36] to code the data. We filtered interesting aspects from the literature by using criteria and sub-criteria derived from theory [36]. Ten concepts were identified for which we defined coding rules [36]. Three researchers coded the articles independently. An initial inter-coder reliability [37] was achieved by calculating the average pairwise percent agreement ($A_0 = 0.815$), Fleiss' Kappa ($\kappa = 0.711$), average pairwise Cohen's Kappa ($\kappa = 0.712$), and Krippendorff's Alpha ($\alpha = 0.711$). Since all values exceed the critical value of $\alpha_{min} = 0.667$, sufficient congruence between the coders can be assumed.

[1] We excluded articles with GDL approach, such as product-service-systems that emphasize a transfer of physical items or temporary access to resources.

4 Conceptual Analysis of the Service Innovation Literature

To present a systematic review of SI literature, we compiled a concept matrix, which is presented in *Table 2*. The matrix is built on two main constructs of ambidexterity theory: exploration and exploitation. The main properties of exploration and exploitation are derived from ambidexterity literature, which are then clustered into three dimensions of service science: potential, process, and outcome [38] to guide the derivation of sub-concepts for analyzing the literature on SI more in-depth.

First, the potential of SI can be grounded in the intention of a new *strategic focus (1)* or *operational leverage (2)*. The potential for exploration is manifested in strategical positioning to gain a sustainable competitive advantage [39] while the exploitation potential is based on leveraging existing services by enhancing efficiency, productivity, or the service itself by changing its value proposition [10]. Second, the process of exploration and exploitation involves the use of resources, such as intellectual resources, organizational resources, technologies, and other assets [44]. Exploration is associated with search and *acquisition of resources (3)*, which are new to an organization or the strategic intent to newly combine internal and external resources to design innovative services [10, 20]. In exploitation the emphasis is on leveraging, adopting, and *reconfiguring resources (4)* that are already prevalent in an organization to capture opportunities in SI and extend existing services in day-to-day work [10]. In this regard, ambidexterity describes the ability of organizations to simultaneously exploit existing resources, and explore new resources for SI [59]. Exploration will require disruption and a *path-breaking approach (5)* by abandoning established services [41] to generate a new value proposition [59]. As opposed to exploration, exploitation is *path-dependent (6)*, since improvements and incremental innovation evolve from existing knowledge and routine-based experience [60]. Through the performance of task and activities new opportunities can emerge that trigger variations of the current services [61]. Third, exploration and exploitation often create divergent outcomes [61]. Even though both are important for the viability of an organization, exploration will rather lead to a *radical innovation (7)* whereas exploitation is associated with *incremental innovations (8)*, e.g. as improvements or extensions of service [30]. The focus of the paper is another concept of the analysis, in which we consider the articles that emphasize the design of an innovative *service system or value proposition (9)*. We also analyze whether *technology (10)* is mentioned in SI.

Some of the analyzed articles propose a framework that structures SI through a configuration of resources and use of organizational capabilities [22, 23, 39, 41, 43–45, 49–52, 55, 56]. An integrated framework that adheres to the opportunities of digital transformation is offered by Lusch & Nambisan [50]. Their framework is based on SDL and treats SI as a collaborative process occurring in actor-to-actor networks, in which individuals interact within a service eco-system [50]. Lusch & Nambisan [50] include service platforms, which enhance the efficiency and effectiveness of SI by de-coupling and re-combing resources to generate new value propositions. Thereby, technology can act as an operand resource—on which an operation is performed to produce an effect [5]—for supportive or enabling purposes and as an operant resource—that acts upon operand resources [5]—by creating novel opportunities for resource integration [50].

Table 2. Conceptual Analysis of Service Innovation Literature

Authors	Exploration				Exploitation				Focus	
	Strategic Focus (1)	Resources Acquisition (3)	Path-Breaking (5)	Radical Innovation (7)	Operational Leverage (2)	Resource Configuration (4)	Path-Dependent (6)	Incremental Innovation (8)	Scope (9)	Technology (10)
Andreassen et al. [39]					x	x	x	x	VP	x
Bessant & Maher [40]	x	x	x	x				x	VP	x
Beverungen et al. [18]	x	x	x	x					SYS	x
Bitner et al. [41]	x	x	x	x				x	VP	x
Breidbach & Maglio [23]					x	x	x	x	SYS	x
Chai et al. [42]					x	x	x	x	VP	x
Den Hertog et al. [43]				x	x	x	x		VP	x
Froehle & Roth [44]	x	x	x	x					VP	x
Helkkula et al. [45]					x	x	x	x	SYS	x
Höckmayr & Roth [46]	x		x	x		x			SYS	x
Jaakkola et al. [47]	x	x	x	x					VP	x
Kindström & Kowalkowski [48]					x	x	x	x	VP	x
Lipusch et al. [49]	x	x	x	x					VP	x
Lusch & Nambisan [50]				x	x	x	x	x	VP	x
Ojasalo & Ojasalo [7]	x	x	x	x					VP	x
Ordanini & Parasuraman [51]		x	x	x				x	VP	x
Ordanini et al. [52]	x	x	x	x					VP	
Patrício et al. [53]	x	x	x	x					VP	x
Patrício et al. [54]	x	x	x	x				x	SYS	x
Rubalcaba et al. [22]	x	x	x	x					VP	x
Russo-Spena & Mele [55]					x	x	x	x	VP	x
Salunke et al. [12]	x			x		x	x		VP	x
Srivastava & Shainesh [56]					x	x	x	x	VP	x
Teixeira et al. [57]	x	x	x	x					SYS	x
Yu & Sangiorgi [58]	x	x	x	x					VP	

Abbreviations: VP = Value Proposition, SYS = Service System

The third element of the framework is value co-creation by integrating existing resources of the service provider and the customer. Russo-Spena & Mele [55] extend the idea of co-creation by introducing the five "Co-s" model, which includes: co-ideation, co-valuation, co-design, co-test, and co-launch. Each "Co-" represents a phase of the innovation process resulting from dynamic and on-going interactions among resources, actions, and a group of actors who are interrelated via a dense network [55].

Other articles propose a linear approach [42, 46, 53, 54, 57] with discrete or consecutive steps or iterative approaches with multiple repetitions of the involved activities [7, 18, 48, 58] for developing innovative services. Ojasalo & Ojasalo [7] develop a lean SI model for iteratively designing a new service through several improvement rounds until it is implemented. Each iterative cycle results in a minimum viable service that is deployed for customer feedback [7]. A different iterative approach is provided by Beverungen et al. [18]. The authors suggest an agile method that incorporates the recombination of internal and external resources to design an innovative service system [18]. The method is based on three phases: service system analysis, service system design, and service system transformation [18]. Each cycle of service system design results in a viable prototype that is either further improved or implemented [18], similar to agile methods in software development.

The conceptual analysis of the literature reveals three insights. First, previous research has mostly focused on either exploration or exploitation of SI. Few papers comprise sub-concepts of both, exploration and exploitation, e.g. in papers that are suggesting to use known and unknown resources for SI. Second, almost all articles consider technology as enabler for the design and delivery of innovative services. Only two articles provide a framework for designing and adopting an innovative service without explicitly addressing technology, neither as resource nor as an opportunity [52]. Third, most articles focus on creating, improving, or managing a value proposition and do not take a service system perspective. Only six articles consider a socio-technical system that comprises people, technologies, and other resources for value-creation.

5 Discussion

5.1 An Integrated View on Service Innovation

From the conceptual analysis we derive three propositions that describe how an integrated view of SI can be achieved. We acknowledge that exploring and exploiting SI are discussed as isolated concepts in the analyzed literature. However, exploration and exploitation should not be treated as a paradoxical tension or sequence of separate phases, but rather as continuum [61]. Ambidexterity theory provides a theoretical foundation that is already manifested in new product development [4]. However, exploring and exploiting SI requires a different approach, since services are designed through co-creation of value by multiple actors [23], integrate the knowledge and skills of multiple actors [5], and provide value through (temporary) access to resources (*value-in-use*) instead of acquisition (*value-in-exchange*) [23].

Our conceptual analysis reveals that SI can arise from both, strategic intent and operational activities. The constructive design of an innovative service is an exploratory process, aiming to design a new value proposition [63]. Managers can trigger innovation top-down through exploration by crafting a vision or strategy on how to design new services. Then, the services are implemented in daily work and become efficient [43]. Continuously adapting and improving services by using existing resources, knowledge, and processes refers to exploitation [20]. Employees have the knowledge needed to perform and improve services [22]. Further, they know the preferences of customers, which they can use to develop new services [22]. Employees can also deviate autonomously from a current service offering if they are faced with

unforeseen incidents, such as conflicts, resistance, specific customer requests, or the inability to carry out a service [22]. By adapting service to overcome constraints, incremental innovation is achieved, which may increase efficiency, effectiveness, or improve other organizational aspects [59]. Ultimately, strategic decisions for exploration are rooted in an organizations' resources and activities that are exploited [44]. Whereas, exploration promotes an organization's long-term viability, exploitation is indispensable to ensure short-term viability [61], resulting in our first proposition:
Proposition 1: Service innovation is a continuum of exploring new and exploiting current resources, knowledge, and processes.

The conceptual analysis also reveals that technologies create both, opportunities and challenges for SI [10]. On the one hand, this implies that technology and rapid-innovation contexts can generate strong pressure for organizations to introduce innovative changes in a value-proposition [6]. On the other hand, a new value proposition can be explored through deliberately searching and acquiring technological resources, which are then translated as a technological option into a value proposition, e.g. as a self-service [43]. Technology can offer an infrastructure that provides communication, collaboration and/or computing capabilities to support innovation [64] and thus, serve as enabler for service processes [4], e.g. web-based services such as shipment tracking of a postal office or online streaming services. Further, it can be adopted for exploitation by incrementally changing the value-proposition, e.g. through customization, as new ways for customer interaction, or by service extension [43].

While the importance of technology for SI is emphasized in many articles, the role and use of technology is divergent. Often, technology is used as an enabler providing the infrastructure, neglecting the innovation potential that is inherent in technology [4]. The continuous development and launch of new digital and smart technologies, such as smartphones, artificial intelligence, or big data trigger the design of innovative services creating a first-mover-advantage. Therefore, we propose that:
Proposition 2: The acquisition and configuration of technology offers a dual potential, as enabler and as trigger of SI.

SI is mainly studied by focusing on designing or enhancing value propositions, in which consumers participate in the development of service through co-creation of value [5]. However, against the backdrop of digital transformation new perspectives on value creation are needed [46]. A service system's perspective helps to understand SI comprehensively [23]. A service system incorporates resources, e.g. technologies and multiple actors that are connected by a value proposition and shared information [15]. Actors develop innovative services through a collaborative process of sharing and accessing these resources within value networks, i.e. service system configurations [23]. For example, Airbnb has designed a service system by leveraging technology and using the properties of their users as resources to provide customers a new innovative value-proposition. Thereby, Airbnb has become the largest global lodging company and brand today for cheap and efficient traveling without owning any properties [65]. A service system perspective can remedy the disjunction of research in exploring and exploiting SI, constituting our third proposition:
Proposition 3: A service system perspective provides a unified research perspective for service innovation.

5.2 Ontology of Ambidextrous Service Innovation

As exemplified in *Table 3*, most articles emphasize that SI is achieved through the interaction of multiple actors (e.g. client and provider, producer and co-producer, or a network of employees or customers). The process of interaction among multiple actors to co-create value is manifested in recognizable patterns of activities [66]. The analyzed articles provide explicit patterns—presented as frameworks, guidelines, methods, or models—that can be used by actors to explore and exploit SI.

Table 3. Exemplary Definitions of Service Innovation

Author	Definition
Breidbach & Maglio [23]	"Service innovation is service system reconfiguration. […] A service system is composed of multiple entities that interact to cocreate value."
Patricio et al. [54]	"Service innovation can be defined as a new process or service offering that is put into practice by an organization, and is adopted by, and creates value for one or more actors in a service network."
Rubalcaba et al. [22]	"A service innovation can be the outcome of innovation networks in which different agents cooperate to coproduce a service-based innovation result."

In addition to a formalized pattern of SI, the interaction between multiple actors can lead to variation in a value-proposition [66]. Employees can leverage and adapt existing services autonomously or as co-creators of value in cooperation with customers, e.g. to remedy inefficiencies in service provision [22]. Further, employees provide knowledge to elicit the formalized design of a new service. This interdependency can be described as generative mechanisms of adapting existing and experimenting with new patterns of activity to achieve SI. The involvement of multiple actors, the presence of recognizable patterns, and the generative mechanisms of exploration and exploitation mirror the concept of routines [66]. Routines are defined as "recognizable patterns of interdependent actions, carried out by multiple actors" [13]. Hence, SI and routines share the same ontological basis [66]. Further, theories to solve the dilemma of simultaneous competing pressures for exploration and exploitation activities have already been framed in routines literature [6]. From routines' theory, we derive insights on the reinforcing processes of exploration and exploitation for ambidextrous SI.

Routines are effortful accomplishments that are continuously and endogenously changing [13]. The dynamism of routines stems from two components that are recursively related: ostensive and performative aspects. The ostensive aspects represent the schematic form of a routine, whilst the performative aspects embody the actual enactment of the routine that is carried out by specific people, at specific times, and in specific places [13]. Ostensive and performative aspects form a mutually constitutive relationship, in which the ostensive aspects enable and constrain the desired performance of a routine whereas the performance of a routine creates and recreates the ostensive aspects [13]. Hence, an incremental variation in the performance can lead to a change of the ostensive aspects and consequently alter the overall routine. Analog to routine's theory, we emphasize a constitutive relationship of exploration and exploitation in SI, as illustrated in *Figure 2*.

The exploration of a new value proposition constitutes the basis for providing innovative services efficiently. Exploration enables and constrains exploitation in day-to-day work (1). In turn, multiple actors who exploit SI in their routines help to create and recreate the exploration of innovative services (2). This perception is in line with *proposition 1*, that SI can be described as a continuum of exploring new and exploiting current resources, knowledge, and processes.

In routine theory, external changes, e.g. a new technology, can induce the need to transform a routine by changing the ostensive aspect [67]. A radical, path-breaking, and strategic transformation of a routine in response to changes in an eco-system refers to the exploration of new routine patterns. Employees can also deviate in their performance from a standard routine by integrating external resources in their daily work or through reflective self-monitoring, which both refer to exploitation [13]. Likewise, actors are searching and acquiring technologies to explore new innovative services (3). The effect of technology that is radically new to the organization is mainly disruptive (4). When multiple actors are exploiting SI, they can also reconfigure a service (5) by selecting and integrating existing technology to adapt the value-proposition (6). These findings are in line with *proposition 2*, stating that the acquisition and configuration of technology offers potential for both, as enabler and trigger of SI.

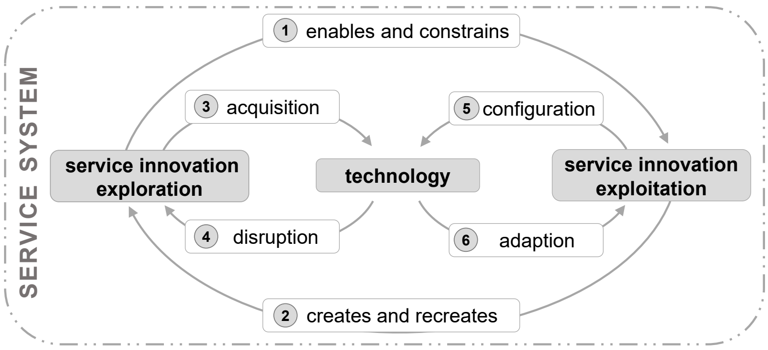


Figure 2. A Systematic View on Ambidextrous Service Innovation

Finally, routines are generative and continuously emerging systems with internal structures and dynamics [68]. A routine represents the interplay between ostensive and performative aspects, the role of artefacts (e.g. technology, rules, and methods), and interaction of multiple actors in shaping these dynamics [68]. By taking a system's perspective, internal dynamics of convergence and divergence among the elements within the system can be observed [68]. Likewise, SI can be viewed as a system (7) that consists of different resources, i.e. people, technology, organizations, and shared information [23]. Within a service system, multiple actors create and co-create mutual value through interaction by sharing and accessing resources [23]. A systematic approach is in line with *proposition 3*, stating that a service system perspective provides a unified research perspective for SI. The findings are summarized in a systematic and unified framework of ambidextrous SI. The framework provides an ample perspective on SI that is characterized by endogenous change. Thereby, the identified gap between distinct strands of research in SI is remedied.

6 Conclusion

Digital transformation provides opportunities for developing innovative services, but also challenges organizations to accelerate innovation cycles of exploring and exploiting new services. The capacity for exploring innovative services while simultaneously exploiting them efficiently is defined as organizational ambidexterity.

Based on a systematic literature review, this paper offers two main contributions. First, despite the need for organizational ambidexterity in practice, we identify that literature on SI consists of two strands of research focusing either on exploration or exploitation. Second, a framework for ambidextrous SI is provided that remedies this disjunction of exploration and exploitation in research. The framework is based on the dynamism and generative mechanisms of routines by adopting the perspective of a mutually constitutive relationship between exploring new and exploiting existing resources, activities, and knowledge. Thus, exploration of innovative services enables and constrains exploitation in day-to-day work. In turn, SI that is exploited in day-to-day work of employees can create and recreate the exploration of innovative services. The framework provides a systematic approach that incorporates technology as an opportunity for designing new and improving existing services. We develop a systematic view on ambidextrous SI that provides a coherent perspective on SI in research and adheres to the demands of organizations, which need to accelerate innovation cycles of exploring and exploiting services simultaneously.

Limitations of this study refer to the coverage of literature that is representative instead of exhaustive. In fact, we only considered peer-reviewed journals and conference proceedings that were published in English. Further, even though inter-coder reliability surpassed the critical value, we cannot assure that other researchers might come to a different assessment of the concepts.

Researchers can use the framework as impetus or guidance to study SI rather as a continuum instead of two distinct concepts. We are eager to see whether our findings can be extended to a broader set of literature and if there are synergistic effects between exploration and exploitation. Our framework can serve as guideline for organizations that can use, test, and assess our theoretical findings in different SI scenarios. Further, we would like to encourage future research to develop theories and methods for ambidextrous SI that take a system's perspective in order to meet organizational demands in a digitally transforming service eco-system.

References

1. Morakanyane, R., Grace, A.A., O'Reilly, P.: Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature (2017)
2. Berghaus, S., Back, A.: Stages in Digital Business Transformation: Results of an Empirical Maturity Study. In: 10th Mediterranean Conference on Information Systems (2016)
3. Hadlington, L., Scase, M.O.: End-user Frustrations and Failures in Digital Technology: Exploring the Role of Fear of Missing Out, Internet Addiction and Personality. Heliyon, vol. 4, e00872 (2018)

4. Ryu, H.S., Lee, J.N.: Understanding the Role of Technology in Service Innovation: Comparison of Three Theoretical Perspectives. Information & Management, vol. 55, pp. 294–307 (2018)
5. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. Journal of Marketing, vol. 68, pp. 1–17 (2004)
6. D'Adderio, L.: The Replication Dilemma Unravelled: How Organizations Enact Multiple Goals in Routine Transfer. Organization Science, vol. 25, 1325–1350 (2014)
7. Ojasalo, J., Ojasalo, K.: Lean Service Innovation. Service Science, vol. 10, 25–39 (2018)
8. Röglinger, M., Schwindenhammer, L., Stelzl, K.: How to Put Organizational Ambidexterity into Practice-Towards a Maturity Model. International Conference on Business Process Management, vol. 16, 194-210 (2018)
9. Witell, L., Snyder, H., Gustafsson, A., Fombelle, P., Kristensson, P.: Defining Service Innovation: A Review and Synthesis. Journal of Business Research, vol. 69, 2863–2872 (2016)
10. O'Reilly III, C.A., Tushman, M.L.: Ambidexterity as a Dynamic Capability: Resolving the Innovator's Dilemma. Research in Organizational Behavior, vol. 28, 185–206 (2008)
11. Patricio, L., Gustafsson, A. and Fisk, R.: Upframing Service Design and Innovation for Research Impact. Sage Publications: Los Angeles, CA (2018)
12. Salunke, S., Weerawardena, J., McColl-Kennedy, J.R.: Competing Through Service Innovation: The Role of Bricolage and Entrepreneurship in Project-Oriented Firms. Journal of Business Research, vol. 66, 1085–1097 (2013)
13. Feldman, M.S., Pentland, B.T.: Reconceptualizing Organizational Routines as a Source of Flexibility and Change. Administrative Science Quarterly, vol. 48, 94–118 (2003)
14. Maglio, P.P., Spohrer, J.: Fundamentals of service science. Journal of the academy of marketing science 36, 18–20 (2008)
15. Vargo, S.L., Maglio, P.P., Akaka, M.A.: On value and value co-creation: A service systems and service logic perspective. European management journal 26, 145–152 (2008)
16. Maglio, P.P, Vargo, S.L, Caswell, N., Spohrer, J.: The Service System is the Basic Abstraction of Service Science. Information Systems and e-Business Management 7, 395–406 (2009)
17. Pöppelbuß, J., Plattfaut, R., Ortbach, K., Malsbender, A., Voigt, M., Niehaves, B., Becker, J.: Service Innovation Capability: Proposing a New Framework. 11th Federated Conference on Computer Science and Information Systems, 545–551 (2011)
18. Beverungen, D., Lüttenberg, H., Wolf, V.: Recombinant Service Systems Engineering. Business & Information Systems Engineering, vol. 60, 1–15 (2017)
19. Böhmann, T., Leimeister, J.M., Möslein, K.: Service Systems Engineering. Business & Information Systems Engineering, vol. 6, 73–79 (2014)
20. Menor, L.J., Tatikonda, M.V., Sampson, S.E.: New Service Development: Areas for Exploitation and Exploration. Journal of Operations Management, vol. 20, 135–157 (2002)
21. Lusch, R.F., Vargo, S.L.: Service-Dominant Logic: Reactions, Reflections and Refinements. Marketing Theory, vol. 6, 281–288 (2006)
22. Rubalcaba, L., Michel, S., Sundbo, J., Brown, S.W., Reynoso, J.: Shaping, Organizing, and Rethinking Service Innovation: A Multidimensional Framework. Journal of Service Management, vol. 23, 696–715 (2012)
23. Breidbach, C.F., Maglio, P.P.: A Service Science Perspective on the Role of ICT in Service Innovation. 23rd European Conference on Information Systems (2015)
24. Gallouj, F., Savona, M.: Innovation in Services: A Review of the Debate and a Research Agenda. Journal of Evolutionary Economics, vol. 19, 149–172 (2009)
25. O'Reilly III, C.A., Tushman, M.L.: Organizational Ambidexterity: Past, Present, and Future. Academy of Management Perspectives, vol. 27, 324–338 (2013)

26. Duncan, R.B.: The Ambidextrous Organization: Designing Dual Structures for Innovation. The Management of Organization, vol. 1, 167–188 (1976)
27. Raisch, S., Birkinshaw, J., Probst, G., Tushman, M.L.: Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. Organization Science vol. 20, 685–695 (2009)
28. March, J.G.: Exploration and Exploitation in Organizational Learning. Organization Science, vol. 2, 71–87 (1991)
29. Gibson, C.B., Birkinshaw, J.: The Antecedents, Consequences, and Mediating Role of Organizational Ambidexterity. Academy of Management Journal, vol. 47, 209–226 (2004)
30. Tushman, M.L., O'Reilly III, C.A.: Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change. California Management Review, vol. 38, 8–29 (1996)
31. Levinthal, D.A., March, J.G.: The Myopia of Learning. Strategic Management Journal, vol. 14, 95–112 (1993)
32. Singh, A., Hess, T.: How Chief Digital Officers Promote the Digital Transformation of their Companies. MIS Quarterly Executive, vol. 16 (2017)
33. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 8–23 (2002)
34. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., others: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. 9th European Conference on Information Systems, 2206–2217 (2009)
35. Cooper, H.M.: Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. Knowledge in Society, vol. 1, 104–126 (1988)
36. Mayring, P.: Qualitative Inhaltsanalyse. In: Mey, G., Mruck, K. (eds.) Handbuch Qualitative Forschung in der Psychologie, pp. 601–613. VS Verlag für Sozialwissenschaften, Wiesbaden (2010)
37. Krippendorff, K.: Content Analysis. An Introduction to its Methodology. SAGE Publications, CA (2013)
38. Hilke, W.: Grundprobleme und Entwicklungstendenzen des Dienstleistungs-Marketing. In: Dienstleistungs-Marketing, pp. 5–44. Springer, Heidelberg (1989)
39. Andreassen, T.W., Kristensson, P., Lervik-Olsen, L., Parasuraman, A., McColl-Kennedy, J.R., Edvardsson, B., Colurcio, M.: Linking Service Design to Value Creation and Service Research. Journal of Service Management, vol. 27, 21–29 (2016)
40. Bessant, J., Maher, L.: Developing Radical Service Innovations in Healthcare—The Role of Design Methods. International Journal of Innovation Management, vol. 13, 555–568 (2009)
41. Bitner, M.J., Ostrom, A.L., Morgan, F.N.: Service Blueprinting: A Practical Technique for Service Innovation. California Management Review, vol. 50, 66–94 (2008)
42. Chai, K.-H., Zhang, J., Tan, K.-C.: A TRIZ-Based Method for New Service Design. Journal of Service Research, vol. 8, 48–66 (2005)
43. Den Hertog, P., van der Aa, W., Jong, M.W.: Capabilities for Managing Service Innovation: Towards a Conceptual Framework. J. of Service Management, vol. 21, 490–514 (2010)
44. Froehle, C.M., Roth, A.V.: A Resource-Process Framework of New Service Development. Production and Operations Management, vol. 16, 169–188 (2007)
45. Helkkula, A., Kowalkowski, C., Tronvoll, B.: Archetypes of Service Innovation: Implications for Value Cocreation. Journal of Service Research, vol. 21, 284–301 (2018)
46. Höckmayr, B.S., Roth, A.: Design of a Method for Service Systems Engineering in the Digital Age, 38th International Conference on Information Systems (2017)
47. Jaakkola, E., Meiren, T., Witell, L., Edvardsson, B., Schäfer, A., Reynoso, J., Sebastiani, R., Weitlaner, D.: Does One Size Fit All? New Service Development Across Different Types of Services. Journal of Service Management, vol. 28, 329–347 (2017)

48. Kindström, D., Kowalkowski, C.: Development of Industrial Service Offerings: A Process Framework. Journal of Service Management, vol. 20, 156–172 (2009)
49. Lipusch, N., Dellermann, D., Kleinschmidt, S., Ebel, P., Leimeister, J.M.: Towards Developing a Service Innovation Canvas, 38th International Conference on Information Systems (2017)
50. Lusch, R.F., Nambisan, S.: Service Innovation: A Service-Dominant Logic Perspective. MIS Quarterly, vol. 39 (2015)
51. Ordanini, A., Parasuraman, A.: Service Innovation Viewed Through a Service-Dominant Logic Lens: A Conceptual Framework and Empirical Analysis. Journal of Service Research, vol. 14, 3–23 (2011)
52. Ordanini, A., Parasuraman, A., Rubera, G.: When the Recipe is More Important than the Ingredients: A Qualitative Comparative analysis (QCA) of Service Innovation Configurations. Journal of Service Research, vol. 17, 134–149 (2014)
53. Patricio, L., Fisk, R.P., Falcão e Cunha, J.: Designing Multi-Interface Service Experiences: The Service Experience Blueprint. Journal of Service Research, vol. 10, 318–334 (2008)
54. Patricio, L., Fisk, R.P., Falcão e Cunha, J., Constantine, L.: Multilevel Service Design: from Customer Value Constellation to Service Experience Blueprinting. Journal of Service Research, vol. 14, 180–200 (2011)
55. Russo-Spena, T., Mele, C.: "Five Co-s" in Innovating: A Practice-Based View. Journal of Service Management, vol. 23, 527–553 (2012)
56. Srivastava, S., Shainesh, G.: Bridging the Service Divide Through Digitally Enabled Service Innovations: Evidence from Indian Health Care Service Providers. MIS Quaterly, vol. 39 (2015)
57. Teixeira, J.G., Patricio, L., Huang, K.-H., Fisk, R.P., Nóbrega, L., Constantine, L.: The MINDS Method: Integrating Management and Interaction Design Perspectives for Service Design. Journal of Service Research, vol. 20, 240–258 (2017)
58. Yu, E., Sangiorgi, D.: Service Design as an Approach to Implement the Value Cocreation Perspective in New Service Development. Journal of Service Research, vol. 21, 40–58 (2018)
59. Andriopoulos, C., Lewis, M.: Exploitation-Exploration Tensions and Organizational Ambidexterity: Managing Paradoxes of Innovation. Organization Science, vol. 20, 696–717 (2009)
60. He, Z.-L., Wong, P.-K.: Exploration vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis. Organization Science, vol. 15, 481–494 (2004)
61. Lavie, D., Stettner, U., Tushman, M.L.: Exploration and Exploitation Within and Across Organizations. The Academy of Management Annuals, vol. 4, 109–155 (2010)
62. Xie, R., Ling, H., Sun, Y., Zhang, C.: Towards the Ambidexterity of Service Processes: Implications for the Managing of IT Usage and Organization Learning. WHICEB, vol. 6 (2013)
63. Barrett, M., Davidson, E., Prabhu, J., Vargo, S.L.: Service Innovation in the Digital Age: Key Contributions and Future Directions. MIS Quarterly, vol. 39, 135–154 (2015)
64. Nambisan, S.: Digital Entrepreneurship: Toward a Digital Technology Perspective of Entrepreneurship. Entrepreneurship Theory and Practice, vol. 41, 1029–1055 (2017)
65. Varma, A., Jukic, N., Pestek, A., Shultz, C.J., Nestorov, S.: Airbnb: Exciting Innovation or Passing Fad? Tourism Management Perspectives, vol. 20, 228–237 (2016)
66. Pentland, B.T., Yakura, E., Singh, H.: Services, Processes and Routines: Literature Review and Implications. 14th Intern. Conference on Information Resources Management, 12–14 (2011)
67. Leonardi, P.M.: When Flexible Routines Meet Flexible Technologies: Affordance, -Constraint, and the Imbrication of Human and Material Agencies. MIS Quarterly, vol. 35 (2011)
68. D'Adderio, L.: The Performativity of Routines: Theorising the Influence of Artefacts and Distributed Agencies on Routines Dynamics. Research Policy, vol. 37, 769–789 (2008)

Design and success factors of an online solution for cross-pillar pension information

Christopher Krug[1]

[1] FOM University of Applied Sciences, Essen, Germany
mail@christopher-krug.de

Abstract. As the European Commission notes, the importance of individual pension components is increasingly shifting from state provision to private provision, as most European states can only guarantee basic provision for retirement age instead of guaranteeing the continuation of a person's standard of living. This development requires transparent information for citizens on the composition of their pensions, as this is a prerequisite for taking responsibility for their old-age provision themselves. Digital solutions such as apps offer the possibility of bundling the various sources of an old-age provision in one system and showing them to the individual citizen at a glance to make individual pension needs and possibilities of influence understandable. This research-in-progress paper reflects the current state of research in a study that examines the success factors of digital retirement provision information in a pan-European context.

Keywords: Pension information, Digital Transformation, Financial Literacy, Pension Literacy

1 Introduction

Due to a pan-European decline in the level of statutory pension entitlements, the responsibility for adequate old-age provision is shifting from the state to the individual citizen [1-3]. Old-age provision in the member states of the European Union (EU) is based on the three pillars of statutory, company and private old-age provision and thus company and private pension provision become more important [1].

To be able to take on this responsibility, it is essential that individuals understand the development of their old-age provision, so they can find out about their pension needs and act accordingly [1-3]. In most EU Member States, however, there is a lack of an appropriate information medium to enable citizens to do so, as cross-pillar entitlements are difficult for non-professionals to understand due to complex information and effects like money illusion [2-4]. In consequence, this makes a significant contribution to the fact that many citizens are uncertain about their pension situation and do not pay enough attention to it [4-6].

The advantages over traditional media regarding data consolidation, simplification, and personalization offer potential for the digital transformation of pension information to solve this issue. Existing implementations, however, like in Denmark or Sweden,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

focus strongly on country-specific issues and thus cannot be transferred to other countries [3]. Hence, the European Commission (EC) recommends the implementation of an overall digital solution for comprehensive pension information, which considers the different pillars as well as the free movement of workers [1], [7]. In this study, which is still in progress, the success factors of such a solution are evaluated.

2 General State of Research on Pension Information

At present, citizens are confronted with a wide variety of sources of pension information, like previous employers or different EU countries [8]. Figure 1 shows which individual service providers the user is currently facing and what added value is generated by a cross-pillar pension information medium.

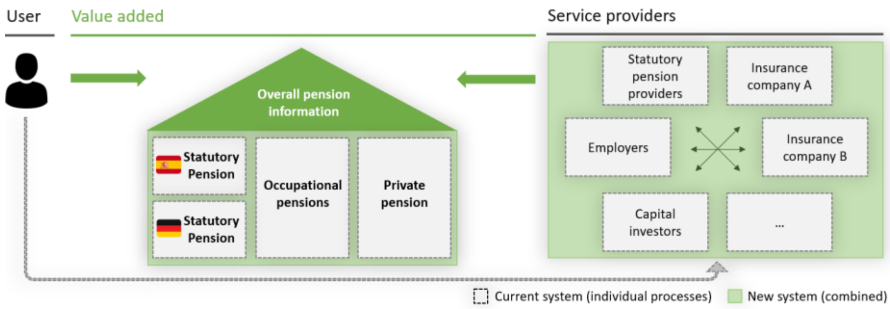


Figure 1. Pension information service system

Increasing the transparency of pension information and bringing together the various pillars is a highly contemporary topic, what is reflected in the number of current working papers and studies on the need for a cross-pillar retirement provision information medium [1], [4], [7-11]. However, the corresponding publications are regularly backed by private sponsors or public organizations. Current scientific research regarding the success factors of pension information, however, relates primarily to the individual pillars like, for example, the statutory pension [3], [12-16].

3 Research Approach

As outlined, the EC, among others, is proposing a digital solution for cross-pillar pension provision for its member states. This research work aims to determine the success factors of such an overall information medium, which is to be considered pan-European and inter-organizational.

In order to create the basis for achieving this objective, an application concept for a cross-pillar information solution has been developed. Since there is hardly any existing literature that specifically addresses the requirements of cross-pillar pension information, the concept is mainly derived by results of expert interviews, in addition to a literature review and a comparison of existing solutions in Denmark and Sweden.

To validate that the concept meets the requirements of the experts, it was prototypically implemented mainly via the application framework Angular and provided via a public domain, so the experts were able to evaluate the implemented solution in a second survey. Table 1 shows the framework data of both surveys.

Expert interviews	Phase 1 – Requirements	Phase 2 – Validation
Number of experts	12	7 (all of them also participated in phase 1)
Heterogeneity	High; experts from all three pension pillars as well as politics and science	Medium; experts from sciences and the second and third pension pillar
Structuring	Semi-structured	Structured
Cognitive interest	Attitudes and information	Information
Role of the interviewer	Active	Electronic, non-personal conduct of the survey
Interview situation	Symmetrical	Asymmetric (in favor of the interviewer)

Table 1. General data of the expert interviews

The second phase of expert interviews also serves as a pre-test before the conduct of end-user-focused experiments. According to the Design Science Research (DSR) approach of Hevner et al., the application design is broken down into several artifacts to strengthen the significance of this prospective end-user evaluation [17]. The findings of this evaluation will be analyzed and iteratively integrated into the concept.

In the first step of end-user evaluation, it is necessary to examine whether the basic application concept also meets the requirements of end users. Due to the complicated information situation and a high demand for a digital solution in Germany, German users are used for the first step. After a successful first evaluation and iterative revision of the concept, an end-user evaluation with other European users will take place. This evaluation is planned to be carried out in Spain and Italy, since the pension composition is different than in Germany, due to the lower relevance of the second pillar.

4 Current State of Results

The developed concept consists of a set of 12 functional, 11 non-functional requirements and a technical design. The functional requirements include: *(1) registration & identification; (2) overview of pension claims; (3) interfaces to pension providers; (4) a detailed view of pension modules; (5) merging claims for planning purposes; (6) input of forecast variables; (7) insight into personal data; (8) input of personal data changes; (9) further information insight; (10) contact options; (11) monitoring; (12) demo access*. The non-functional requirements mainly cover the framework conditions of the solution, such as access with as few barriers as possible and high usability. The technical design covers a first system design in the form of a web application with three-layer architecture and drafts for each of the layers. Specifically, these are a) an information architecture based on a mobile-first approach for the presentation layer, b) process designs for the application layer and c) a design for the data layer, including an entity relationship diagram for all pension-relevant data.

The expert interviews and the pretest carried out via a prototype implementation of the concept confirmed that the concept fulfils the experts' requirements. The prototype included all non-functional requirements as well as the functional requirements 1-10 on the presentation layer. Furthermore, basic functions were implemented on the application layer, e.g. for adding/changing/deleting pension entitlements. With regard to the interview results, it should be emphasized that the feasibility of cross-pillar information is supported by the providers of all pillars, with most of the experts taking the view that the respective other providers would be opposed to a cross-pillar solution. Also, the obstacle to data protection, for example, is assessed as solvable, insofar as data are only transferred on-demand and end-to-end to the respective user. Automatic data provision is identified as a first, decisive success factor, which requires a uniform identification of the users, e.g. via tax ID or social insurance number, as it exists in most European countries. When asked how the experts would assess the concept without fulfilling this success factor, it was judged to be less effective, but still an improvement over singular information media. This assessment supports the relevance of a cross-pillar information medium, which will be further deepened by the end-user evaluation.

5 Conclusion

This paper presents the basic methodology as well as the main results of the research work so far. Due to the focus of this research-in-progress-paper, the basics, in particular, represent only a small part of the research object, which opens up numerous further research fields. Thus, critical elements for pension information, such as the individual service providers will take on a more significant role in the final research.

The first evaluation showed that the requirements of experts from different pillars could largely be combined in one application, for which initial success factors and challenges could also be identified. In the next steps, these initial findings will be incorporated into the user-centred evaluation at a national and pan-European level as well as into the iterative development of the concept following the DSR approach.

References

1. European Commission: Grünbuch: Angemessene, nachhaltige und sichere europäische Pensions- und Rentensysteme. European Commission, Brussels (2010)
2. Antolin, P., Stewart, F.: Private Pensions and Policy Responses to the Financial and Economic Crisis. OECD Working Papers on Insurance and Private Pensions, No. 36. OECD Publishing, Paris (2009)
3. Atkinson, A., Harrison, D., Messy, F., Yermo, J.: Lesson from National Pensions Communication Campaigns. OECD Working Papers on Finance, Insurance and Private Pensions, No. 18. OECD Publishing, Paris (2012)
4. Haupt, M., Wegner-Wahnschaffe, C.: Trägerübergreifende Altersvorsorgeinformationen – eine unendliche Geschichte? BetrAv Vol. 72(3), 233-240 (2017)
5. Börsch-Supan, A., Bucher-Koenen, T., Coppola, M., Lamla, B.: Savings In Times Of Demographic Change: Lessons From The German Experience. Journal of Economic Surveys Vol. 29(4), 807-829 (2015)

6. Rürup, B., Huchzermeier, D., Böhmer, M., Ehrentraut, O.: Die Zukunft der Altersvorsorge. Handelsblatt Research Institute, Dusseldorf (2014)
7. European Commission: EaSI-PROGRESS: Implementing the pilot stage of the European Tracking Service for pensions, http://ec.europa.eu/social/main.jsp?catId=629&langId=de&callId=510&furtherCalls=yes (Accessed: 12.08.2018)
8. Stevens, Y.: General Evaluation Of The European Communication On Pension Rights. In: European Colloquium: Communication on pension rights in Europe, pp. 71-89. Marc de Block, Brussels (2010)
9. Spendler, S., Morgenstern, K.: Online-Rentenkonto – Deutschland hinkt hinterher. Deutsches Institut für Altersvorsorge, Berlin (2018)
10. Hackethal, A., Bucher-Koenen, T., Kasinger, J., Laudenbach, C: Wie viel Rente wirst du bekommen? Pilotstudie zur Einführung einer Renteninformationsplattform in Deutschland. Working Paper, SAFE - Sustainable Architecture for Finance in Europe (2018)
11. Haupt, M., Sesselmeier, W., Yollu-Tok, A.: Ungeklärte Diskrepanzen. Der verhaltensökonomische Blick auf die Altersvorsorge. In: Betzelt, S., Bode, I. (eds.) Angst im neuen Wohlfahrtsstaat 2018. HWR Berlin Forschung, Bd. 64, pp. 339–368. Nomos Verlag, Berlin (2018)
12. Kritzer, B. E., Smith, B. A.: Public Pension Statements in Selected Countries: A Comparison. Social Security Bulletin Vol. 76(1), 27-56 (2016)
13. Fleishman-Mayer, L., Hung, A., Yoong, J., Clift, J., Tassot, C.: Designing Better Pension Benefits Statements. Working Paper, RAND Labor & Population (2013)
14. OECD: Pensions at a Glance 2017: OECD and G20 Indicators. OECD Publishing, Paris (2017)
15. Ebbinghaus, B.: The Privatization and Marketization of Pensions in Europe: A Double Transformation Facing the Crisis. European Policy Analysis Vol. 1(1), 56-73 (2015).
16. Dolls, M., Doerrenber, P., Peichl, A., Stichnoth, H.: Do Savings Increase in Response to Salient Information about Retirement and Expected Pensions? Working Paper, National Bureau of Economic Research (2016)
17. Hevner, A., March, S., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly Vol. 28(1)., 75-105 (2004)

Track 7:

IT-Management und -Strategie

Track Chairs:

Prof. Dr. Nils Urbach
Universität Bayreuth

Prof. Dr. Paul Drews
Leuphana Universität Lüneburg

A Frugal Support Structure for New Software Implementations in SMEs

Nivedita Agarwal[1], Martin Schymanietz[1], Albrecht Fritzsche[1]

[1] Friedrich-Alexander University Erlangen-Nürnberg, Institute of Information Systems, Nuremberg, Germany
{nivedita.agarwal,martin.schymanietz,albrecht.fritzsche}@fau.de

Abstract. During software implementations, budgetary and human resource constraints often make it difficult for small and medium sized enterprises (SMEs) to provide and maintain the required support. To overcome these constraints, this study describes a frugal support structure (FSS) to orchestrate available resources and to involve users as suppliers and co-creators of contextualized information. The FSS is conceptualized as a system that enables interaction and collaboration between the actors involved by using extant communication infrastructure wherever possible, systematizing and centralizing knowledge created and ensuring overall resource and time efficiency. Adopting a design science research process, development of the FSS combines a literature review and practical insights. Evaluating the challenges and benefits of FSS, the findings indicate that user involvement is necessary not only for contextualized and accessible support but to make support structures more frugal and sustainable in the long term.

Keywords: Software implementations, frugal, design science, user involvement, SMEs

1 Introduction

Software implementations have always interested IS scholars because of their high-risk and high-reward characteristics [1]. When introducing new software solutions, organizations commonly invest significant resources in change management practices and support structures to facilitate employees' gradual transition to the new business processes [2]. Unlike larger incumbents, resource-constrained small and medium-sized enterprises (SMEs) may find it challenging to support such transitions [6], as evidenced by the low success rates and sluggish adoption of new software in SMEs [1-2], [6].

Previous research on software implementations shows that post-project measures such as training, online support and IT help desks account for almost 90% of the total cost of implementation [6], [10]. Limited financial capacity, low human capital and fragmented governance structures [15] make it more difficult for SMEs to offer and sustain these employee supports over time. However, as the literature confirms, the importance of these support structures for successful software implementation [1], [6], SMEs must balance the costs against the benefits of the new software, making it important to explore more resource-efficient and sustainable alternative supports.

Although user involvement is widely acknowledged as a critical factor for successful software implementation [17], it is rarely a feature of existing support structures [6], [19-20]. The limited attempts to facilitate user involvement have emphasized social structures such as advice networks and peer-to-peer collaboration [6], [19-20], which are said to provide contextualized information and better accessibility than traditional support structures (TSS). As this is a relatively new area of IS research, there are no clear design principles and little consensus in relation to the benefits of these social support structures in different organizational contexts [6]. The present study describes the design of one such structure to enhance user involvement, with particular reference to new software implementation in SMEs.

SMEs commonly face financial and human resource constraints, and a frugal approach is necessary to ensure the efficient utilization of available resources. The frugal approach involves developing cost-effective and accessible solutions by making creative use of resources at hand [21]. We argue here that involving users as consumers, suppliers and co-creators of information will be more cost- and time-efficient. To that end, the present study describes the design of frugal social support structures for SMEs to facilitate creative orchestration of available resources for higher benefits. In designing a frugal support structure (FSS) for SMEs and evaluating its benefits as compared to TSS, we addressed the following research question:

How can frugal support structures orchestrate available resources and influence user adoption within SMEs to overcome specific organizational constraints during IS implementation?

To design and assess the proposed FSS, we adopted the conceptual lens of service-dominant (S-D) logic, which specifies principles for the creative orchestration of interactional resources, including tangible (technological) and intangible resources (knowledge, skills and competencies) by structuring, bundling or leveraging these for competitive advantage [40]. Adopting a design science research (DSR) approach, the subsequent empirical study conceptualized the proposed FSS in terms of S-D logic, followed by ongoing evaluation of the effects on user adoption as compared to TSS. The study was conducted in collaboration with a German SME from the IT sector currently undertaking multiple new software implementations. Drawing on links between the principles of S-D logic and the frugal approach, we focused on user involvement and co-creation. As well as contextualizing S-D logic in a resource-constrained setting, the study describes design guidelines for FSS development and provides evidence of the tangible benefits of a frugal approach to software implementation.

The remainder of this article is structured as follows. After introducing the theoretical foundations of S-D logic, we go on to discuss the frugal approach. We then describe our methodology, which is based on design science research (DSR), along with insights gathered and iterative development and demonstration of the FSS. Following an explanation of the evaluation phase, the article ends with contributions and conclusions.

2 Background

2.1 S-D logic

In contrast to goods-dominant (G-D) logic, which is grounded in a "push" philosophy, S-D logic is based on a user-centric "pull" philosophy and focuses on value co-creation [23], emphasizing process and the exchange of services [24]. On this view, value is co-created by combining the unique resources (e.g., knowledge, skills) of the actors involved (e.g., employees, partners, suppliers, firms, customers) [25], and traditional goods become a mere vehicle for the exchange of value [26].

S-D logic combines tangible and intangible resources that are internal or external to the actors, referred to as *interactional* resources [23]. Technology and knowledge are important types of interactional resource; while technology provides necessary infrastructure, knowledge and specialized skills serve as the fundamental unit of exchange [26]. Actors orchestrate these interactional resources to help each other [26]. S-D logic has four meta-theoretical foundations: actor-to-actor networks, resource liquefaction, resource density and resource integration [24]. Altogether, they provide a strong conceptual basis to address the increasing challenges of systems design and implementation in the digital economy [27]. An actor-to-actor network includes all the relevant actors serving variously as producers, suppliers or users as potential co-creators of value. Resource liquefaction is the decoupling of information from users or technologies and enabling information sharing. Resource density specifies the mobilization of resources in terms of space, time and actors; density is optimized when contextually relevant information is shared in the most effective and efficient way. Resource integration is based on the fundamental idea that resources are less useful in isolation and must be combined with other resources to yield higher value [24].

2.2 The Frugal Approach

Derived from the Latin word *frugalis*, the concept of *frugal* has local equivalents around the globe, such as DIY in the US, Jugaad in India, Zizhu in China, Jua Kali in Africa and système d in France [29]. Scholars have defined frugal innovation as a bottom-up approach to innovation that creates accessible and affordable solutions for resource-constrained customers [30]. Beyond mere de-featuring or remodeling of existing solutions, frugal is a problem-solving approach to innovation whose underlying principle is to "do more with less" [34] – that is, to develop solutions with a higher performance-to-cost ratio [36]. Based on a clean-slate approach, frugal innovation involves re-designing the whole development process to eliminate unnecessary costs, yielding resourceful and easy-to-use solutions [30], [34], [36]. These simple, low-cost, high-benefit, local-focused, scalable, mass-market solutions are designed for the harsh conditions that prevail in emerging markets, responding to the unique needs of customers living in resource-constrained areas. As well as identifying core values and avoiding needless costs, the frugal approach is driven by the concept of *inclusivity* – involving users as suppliers – which is widely discussed in the literature as a means of overcoming particular local constraints [31]. In the context of information systems, frugality is defined as a "*system which is developed and deployed with minimum resources to meet the pre-eminent goal of the client*" [42]. Watson et al. [42] highlighted ubiquity, uniqueness, unison and universality as the four drivers of frugal information systems.

A parallel is often drawn between "bricolage" theory [33] and frugal innovation because of the shared focus on the efficient utilization of resources [40]. While the theoretical foundations of frugal innovation remain contested, the approach is gaining momentum in both developing and developed regions. In developing countries, cost-effective and accessible innovations are needed to overcome extreme conditions and existing resource constraints [36]. In developed countries, companies are turning to frugal innovation in response to changing environmental conditions that include resource scarcities and changing demographics. In light of the close link between the frugal approach, bricolage and inclusivity, we argue here that S-D logic – and especially the fundamental principles of resource density and integration – are highly relevant in resource-constrained contexts. On that basis, the proposed FSS employs S-D logic to design supports for new software implementations in resource-constrained SMEs.

3 Design Science Research

In designing the proposed FSS, we followed the approach of Peffers et al. (2006), who described a design science research process (DSRP) model for iterative building and evaluation of the given artefact [35]. To begin, DSRP specifies the problem by combining insights from a literature review and practical experiences. Having identified the problem, the requirements and objectives of the proposed artefact are then specified. By means of an iterative process, the design is further developed, demonstrated and evaluated in the given context. Finally, the built artefact is communicated to the wider world.

In the present case, this approach was adopted to design and implement an FSS for a German SME (around 250 employees) in the IT sector, which was undertaking several new software implementations to achieve a common cloud-based IT infrastructure. The systems to be implemented included a new intranet, a novel travel management tool, MS Office 365 and a mobile device management system. While some of these were sourced from external vendors, some were developed in-house and required different support structures. Given the challenges of supporting the new IT strategy with limited resources, the company was looking for an integrated and efficient support structure, so providing an appropriate context for our research.

3.1 Problem Identification and Motivation

Literature Review

In a literature review of support structures for software implementations, peer-reviewed journal articles published after the year 2000 were searched across three databases (AIS, EBSCOhost and Science Direct) using the following search stream: ("software introduction" OR "software launch" OR "project launch" OR "project implementation"

OR "Enterprise Resource Planning System*" OR “Enterprise System*" OR “ERP” OR "change management") AND ("user adaptation" OR "effective use" OR “post-implementation” OR “pre-implementation” OR “support” OR “job stress” OR “impact on employees” OR “satisfaction" OR “knowledge management” OR “shakedown phase” OR “learning” OR “acquisition” OR “sense-making” OR "community platform" OR “communities” OR "software agent" OR “collaboration”). Appendix 1 summarizes the 33 articles selected to capture the benefits and challenges of existing support structures as a concept matrix [36]. Largely influenced by ERP implementations, the extant IS literature discusses five primary types of support structure (see Appendix 1), and the benefits and challenges of each are discussed next.

Training. In different forms (e.g., one-on-one, group, online), training is the most widely adopted support structure for new software implementation. Hands-on experience [37-38] and didactic knowledge transfer during training have been shown to impact significantly on user adoption [28]. While recent research links the convenience of online training to higher user acceptance [38-39], it also highlights some associated challenges [6], which include lack of contextual information, high cost and being time bound.

Online Support. Offering real-time support for users, this includes access to manuals, help files and, in some cases, online chat with technical advisors. Users can retrieve information regardless of location or time. However, like any knowledge management database, this type of support must be updated regularly. As well as being human resource-intensive, it lacks the contextual information that users need [6].

Peer-to-peer/Advice Network. This approach involves direct interaction among users with the goal of seeking or giving advice [6]. These networks commonly involve informal exchanges between actors [22], often fellow employees; benefits include ease of access, prompt responses and context-specific knowledge, which means that the information received is often more understandable and more readily applied [6].

Top Management Support. Usually seen as a more intangible support, this relates to resource provision and making the new software visible within the organization [11, 13-14, 16, 18]. Open communication and alignment of software implementation with company objectives have been shown to impact positively on user adoption [11]. However, because of its intangible nature, this kind of support is difficult to realize and is often overlooked [1], [7-9]

IT Helpdesk. Like online support, an IT helpdesk aims to provide generic support to users by facilitating access to manuals or other reference material as an intermediary between software provider and user [22]. Offering mainly technical assistance, this approach fails to provide contextual information; it is also time-consuming, as users must raise an IT ticket each time and then wait for assistance.

Aside from a consistent lack of user involvement, the literature review reveals distinct challenges for TSS in terms of timeliness, resource intensiveness and contextualization of information. While some recent research has sought to demonstrate the benefits of user involvement in the form of advice networks, further exploration is needed. The review also shows that support for new software is currently offered primarily during the pre-implementation phase. In contrast to the strong emphasis on training and top management support, there has been little exploration of other forms of support to promote user involvement, such as peer-to-peer or advice networks. Although it has been identified as a critical factor for successful implementation, the issue of user involvement has been largely ignored. The review also reveals a lack of comparative studies of these support structures in terms of their individual cost and time effectiveness and their impact on user adoption.

Expert interviews

To integrate insights from the literature with practical understanding and for increased relevance [5], expert interviews and discussions were conducted around software implementations at the selected German SME [45]. The company, which was involved in multiple IT rollouts, was selected through convenience sampling [46]. SMEs are of great relevance to the present study for a number of reasons. First, SMEs are of great importance to Germany's economy, accounting for 35.3% of the total revenue of German firms and employing 58.3% of workers who pay social insurance contributions [43]. Additionally, SMEs often lack the financial and human resources of large firms and have different governance structures [16, 44]. To assess the SME's requirements, the authors conducted three detailed interviews and held discussions with key actors, including IT support and project management teams. We also participated in kick-off workshops and introductory sessions for the new software systems. In addition to monthly discussion meetings with the project manager, three further interviews were conducted with other relevant actors. All the interviews were recorded, transcribed and coded with QDA software. The research partner also granted us access to their internal portals to observe user adoption.

The SME started its transformation journey by rolling out off-the-shelf ERP software. Purchased from an external vendor, this software streamlines and integrates business processes across departments such as finance, marketing and sales. As an off-the-shelf product, there is little or no possibility of customization, and the provider offers only standard tutorials and limited training. Thirty lead users (department heads) were selected by the company for initial training, and they were then expected to support further roll-out in their respective departments. Within the company, one project manager was responsible for the entire roll-out and was the single point of contact for end users. Communication between lead users and the project manager mainly involved email exchanges and personal meetings. In the absence of a common communication platform or forum, the project manager often spent a lot of time answering repetitive questions from both lead and end users. The project manager was supported by an IT service desk managed by one full-time employee. For every support requirement, users had to raise a ticket and wait for manual confirmation of the estimated time. As well as being time-consuming, this process was inefficient in terms of utilization of available resources.

In parallel to ERP implementation and the Office 365 initiative, the company also introduced new software for travel and mobile device management. Supported by the same project team, the company continued to struggle to offer adequate support to end users. Interviews with the project team and management highlighted budgetary and human resource constraints and the lack of centralized knowledge management initiatives. While acknowledging the need for user involvement and bi-directional communication, the interviewees expressed resistance to doing so because of the anticipated effort and cost, highlighting the need for a frugal solution that makes best use of extant communication infrastructure wherever possible.

Requirements of a Frugal Support System

Based on the literature review and the practical insights gleaned from the case study, we identified eight main requirements for FSS development on the basis of existing infrastructure. They can be broadly categorized as either conceptual or operational. The conceptual requirements are based on information gathered from the literature on frugal innovation and S-D logic while the operational requirements derive from the research setting.

Conceptual requirements:

1. To facilitate user involvement, a frugal support structure needs to identify the relevant actors.
   Research on frugal innovation emphasizes co-creation with end users and bottom-up development of solutions [32]. This aligns with the S-D logic perspective, which focuses on value co-creation involving all of the actors involved [25]. For that reason, it is important to identify all relevant actors, including end users, lead users and project and IT teams.
2. To increase the interaction between actors, a frugal support structure must establish an actor-to-actor network.
   According to S-D logic, actor-to-actor networks blur traditional provider-seeker relationships and enable co-creation [24]. It is crucial to activate or develop a network that enables the identified actors to exchange their knowledge in order to co-create value for themselves and others. This feature of peer-to-peer sharing is expected to enable the required flexibility and will provide contextual information to FSS.
3. To be cost-effective, a frugal support structure must be based on a universal and modular architecture that facilitates efficient utilization of interactional resources.
   Frugal innovation is about developing cost-effective solutions [36] and a universal platform to overcome the friction of technology incompatibilities [42]. In software contexts, a modular architecture enables the dynamic combination, replacement or replication of available resources.
4. A frugal support structure needs to mobilize and orchestrate resources to increase efficient utilization.
   According to both frugality and S-D logic, it is important to be resourceful as well as liquefying or mobilizing resources. Apart from acquiring and bundling of resources, the literature indicates that orchestration also involves leveraging resources through strategies such as mobilizing or modularizing [4]. This avoids additional costs for setting up isolated applications and ensures the best combination of mobilized resources for a given situation.

Operational requirements:

5. A frugal support structure should combine or provide a common interface for all new software introductions.
   Our interactions with the research partners confirmed that support for different software implementations should be integrated and centralized, as a common interface and consistent navigation makes access smooth and easy to understand. One-time authentication should be implemented to prevent additional hurdles.
6. For value co-creation, a frugal support structure should provide tools for collaboration, as well as distinct but connected communication channels.
7. For knowledge management, users need to be able to easily share and store the information generated.
   Collaboration tools and communication channels must be provided to facilitate both value co-creation and knowledge management. Facilities for sharing and saving user-generated information will co-create a common knowledge base that can be (re)used in certain scenarios.
8. To be time-efficient, a frugal support structure needs to be accessible anytime and anywhere.
   The structure needs to be lightweight and accessible regardless of time and location. It should be able to take account of the urgency level and offer the required support within a specified time.

3.2 Objectives of the proposed solution

Based on the literature review, frequently experienced challenges for TSS include a lack of contextual information, high costs, and time and resource inefficiencies (e.g., in relation to training, online support, IT helpdesk). Advice and peer networks, though better, are not necessarily seen as formal support structures. Based on the frugal and S-D logic approaches, the following objectives were formulated for the proposed FSS to overcome the limitations of TSS and to encourage user involvement. The objectives summarized in Table 1 reflect requirements identified from the literature and from practical insights.

1. Enable interaction between multiple actors through collaborative tools/communication channels
   The solution should enable the establishment of a network that includes all relevant actors and facilitates interaction between them – for instance, by means of a platform that supports both synchronous and asynchronous communication.
2. Provide a centralized and structured self-sustaining system
   The solution should provide a structure that helps actors to find what they are searching for, based on the establishment of communities and/or groups focusing on similar topics. These communities/groups should be managed by the actors themselves.

Table 1. Objectives of the intended frugal support structure

	Objective 1	Objective 2	Objective 3	Objective 4
Description	Enable interaction between multiple actors through collaborative tools/ communication channels	Provide a centralized and structured self-sustaining system	Ensure efficient knowledge management	Ensure resource and time efficiency
Underlying characteristics	R1: Facilitate user-involvement R2: Increase interaction between actors R6: Provide tools for collaboration	R5: Single /common Interface R8: Easy access	R7: Tools for knowledge management	R3: Cost-effective & resource efficient R4: Mobilization of resources

3. Ensure efficient knowledge management
   To facilitate the integration of actors' resources, the solution should support knowledge sharing and should be accessible 24/7. The solution should be overarching – that is, it should provide support across an extensive range of use cases within a single structure.
4. Ensure resource and time efficiency
   In line with the concept of frugality, the solution should be based on existing software and should support modular extension for resource efficiency. Additionally, the system should motivate users to provide solutions within a certain time limit.

3.3 Design and Development

The FSS was developed in four recursive iterations. The first design consisted of a draft based on the literature review and the interviews. Subsequently, a clickable mock-up was created using Atlassian software. In the third iteration, that mock-up was ported to the company's intranet portal. Focusing on the ERP implementation, a special group was created on the intranet website, with customized video tutorials and documents. However, the technical limitations of the platform in terms of collaboration tools and communication media meant that opportunities for user involvement were minimal. For that reason, the fourth iteration incorporated the Microsoft (MS) Teams platform to enable the envisioned support structure. As part of the Office 365 bundle, this digital platform provides a common environment for the formation of actor networks and for co-creation. The design requirements were then implemented on the platform.

3.4 Demonstration

The final design of the FSS on the MS Teams platform, which was available to the whole company, is shown in Figure 1.

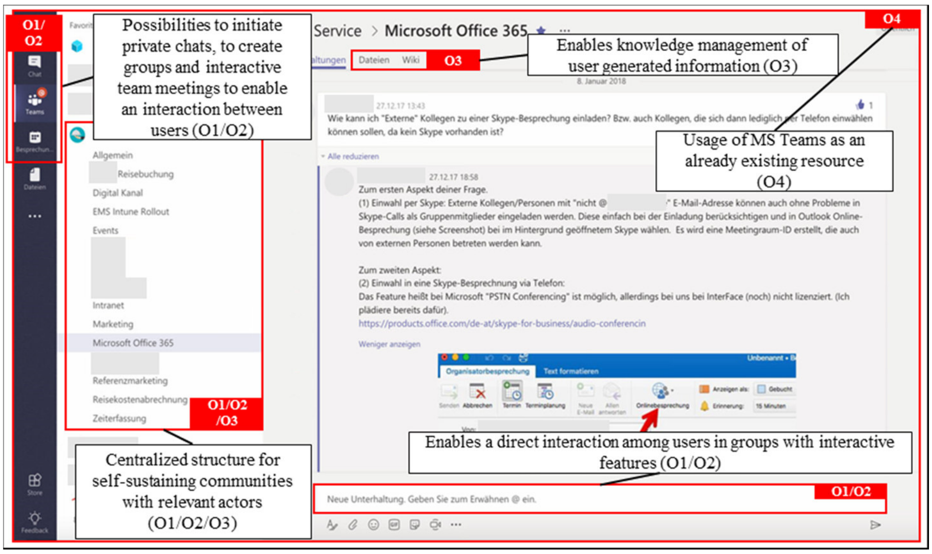


Figure 1. Frugal support structure in MS Teams

3.5 Evaluation and Communication

Evaluation of the FSS was based mainly on the analysis of usage metrics [36] as summarized in Table 2. Analysis of the log files showed that the new FSS was accessed by 163 of the approx. 250 employees from different departments, enabling them to interact and collaborate (Objective 1). Of these, 82 could be regarded as active users of MS Teams and were classified as publicly active, privately active or a combination of both. While only five users were only publicly active (i.e., their posts were accessible to all registered users), 15 employees were solely privately active, using only features like private chat messages to communicate. However, the largest group (62) used both private and public channels to interact with their colleagues. On average, publicly active employees created 34 public posts and 200 private chat messages per user (Objective 1).

The interaction between the different users for support reasons was organized in self-sustaining groups within MS Teams and was used by the employees to ask different questions, concerning the travel booking procedure, mobile device management, MS Office 365, time keeping, intranet, digital services or other general issues (Objectives 2&3). This area was frequently used by 27 users that created a total amount of 244 posts that could be divided into 200 on topic and 44 off topic posts. The employees asked 67 questions in this support structure and 61 of these were answered by other employees. On an average, three posts were created for each of the question asked on the platform.

The analysis of the resource availability showed that 46% of the users just seek information and 32% of them only responded to open questions. Nevertheless, 22% of the active users in the support area took both roles and searched for information as well as pro-vided useful comments on posted problems. While the traditional, ticket-based support clocked a response time of around one week, the frugal support system was able to respond to 90% of the posted questions/problems in less than one week time. Out of which, 73% of questions were answered within one day and 46% in less than two hours, validating a huge reduction of the standard response time through the FSS (Objective 4). However, it emerged that each of the support channels for travel booking, MS Office 365, etc. had one specialist who answered to the maximum number of questions. For example, for travel booking 39% of the posted questions were answered by a single specialist and in case of mobile device management the response from the specialist was as high as 83% of the posted questions.

Table 2. Summary of evaluation

Objective #	Achieved through…
1	•establishment of a wide and active user base •possibilities to communicate on a public and private level
2	•structured groups for different service topics
3	•knowledge sharing among users through their interaction
4	•usage of already established software •increase of response time in case of requests

4 Discussion

To assess the ability of frugal support structures to orchestrate available resources to address specific organizational constraints during IS implementation, the study included a design science research project. This revealed that establishing a frugal support structure can be related to the meta-theoretical foundations of S-D logic. The development of actor-to-actor networks fosters co-creation of value, both dyadic (between information seeker and support channel expert) and at extended network level, where other regular platform users try to help each other [3]. The platform itself supports liquefaction of resources through digital decoupling of single-user knowledge, making this information available to every member of the support structure [24]. High resource density is enabled by response rates that are achievable only by rapid mobilization of actors who can provide the relevant knowledge [4]. Finally, the frugal support structure helps to integrate resources through recombination of existing resources (in this case, IT infrastructure and knowledge), ensuring a level of utility for the whole organization beyond that of isolated resources [24].

The findings confirm that frugal support structures can be systematically developed to support the management of software implementation in SMEs or other settings where there are limited resources for training and change management. At the same time, the present case illustrates that frugal support structures are necessarily diverse and must be designed to ensure that implementation maximizes the potential benefits in each case. Further research on the management of online platforms is needed to devise suitable intervention strategies for increasing traffic and productivity on the platform.

5 Conclusion and Contributions

This article proposes a frugal approach to enhance existing support structures for new software introductions in resource-constrained SMEs. The findings confirm the close links between the frugal approach and S-D logic, where users and their knowledge are the most widely available resource, and value is co-created through integration of the actors' resources. The study confirms that user involvement is crucial, not only in overcoming the challenges of existing support structures but to make support structures more frugal and sustainable in the long run.

To date, the frugal literature has focused on emerging markets and on bottom-up product development approaches, neglecting the developed world, especially in the context of information systems and software implementation. This study is among the first to explore the application of the frugal concept to information systems in developed markets. Using the principles of S-D logic in combination with the frugal criteria of being resourceful and user-driven, the study suggests guidelines for developing a frugal support structure for software implementation. The findings contribute to the S-D logic and frugal innovation literatures. While past research applied S-D logic to the development of efficient customer networks and marketing business solutions [23], the link to frugal innovation contextualizes S-D logic in resource-scarce settings. Along with design principles for developing a frugal support structure, the study clarifies how interactional resources (such as technology) can act as enablers for the development of frugal solutions in developed world settings. Both of these new insights invite further research. From the IS perspective, this study also contributes to the research on software implementation, and in particular to the issue of user involvement, providing evidence of its tangible benefits in the post-implementation process, which remains largely unexplored in the existing literature.

6 Acknowledgement

The research presented in this paper was partially funded by the German Federal Ministry of Education and Research in course of the project ExTEND (www.projekt-extend.de), FKZ 02K14A170ff.

References

1. Liu, G.H.W., Wang, E., Chua, C.: Leveraging Social Capital to Obtain Top Management Support in Complex, Cross-Functional IT Projects. J. Assoc. Inf. Syst. 16:8, pp. 707–737 (2015)

2. Chang, S., Yen, D.C., Huang, S.-M., Hung, P.-Q.: An ERP System Life Cycle-Wide Management and Support Framework for Small- and Medium-Sized Companies. Commun. Assoc. Inf. Syst. 22, pp. 275–294 (2008)
3. Ekman, P., Raggio, R.D., Thompson, S.M.: Service network value co-creation: Defining the roles of the generic actor. Ind. Mark. Manag. 56, pp. 51–62 (2016)
4. Böhmann, T., Leimeister, J., & Möslein, K.M.: Service Systems Engineering. Bus. & Inf. Syst. Eng. 6:22, pp. 73–79 (2014)
5. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Q. 28:1, pp. 75–105 (2004)
6. Sykes, T.A.: Support Structures and Their Impacts on Employee Outcomes: A Longitudinal Field Study of an Enterprise System Implementation. MIS Q. 39: 2, pp. 473–495 (2015)
7. Bingi, P., Sharma, M.K., Godla, J.K.: Critical Issues Affecting an ERP Implementation. Inf. Syst. Manag. 16:3, pp. 7–14 (1999)
8. Iacovou, C.L., Nakatsu, R.: A risk profile of offshore-outsourced development projects. Commun. ACM. 51:6, pp. 89–94 (2008)
9. Dong, L., Neufeld, D., Higgins, C.: Top management support of enterprise systems implementation. J. Inf. Technol. 24:1, pp. 55–80 (2009)
10. Haake, P., Maedche, A., Mueller, B., Lauterbach, J.: The Effect of User Adaptation on the Effective Use of Enterprise Systems. In: Proceedings of the Thirty Sixth International Conference on Information Systems (ICIS), pp. 1–12 (2015)
11. Zhang, Z., Lee, M., Huang, P., Zhang, L., Huang, X.: A framework of ERP systems implementation success in China: An empirical study. Int. J. Prod. Econ. 98:1, pp. 56–80 (2005)
12. Wagner, L.E., Newell, S.: Exploring the Importance of Participation in the Post-Implementation Period of an ES Project: A Neglected Area. J. Assoc. Inf. Syst. 8:10, pp. 508–524 (2007)
13. Sun, A.Y.T., Yazdani, A., Overend, J.D.: Achievement assessment for enterprise resource planning (ERP) system implementations based on critical success factors (CSFs). Int. J. Prod. Econ. 98:2, pp. 189–203 (2005)
14. Nah, F.F.-H., Zuckweiler, K.M., Lau, J.: ERP implementation: Chief information officers' perceptions of critical success factors. Int. J. Hum.-Comput. Interact. 16:1, pp. 5–22 (2003)
15. Parry, S., Jones, R., Rowley, J., Kupiec-Teahan, B.: Marketing for survival: a comparative case study of SME software firms. J. Small Bus. Enterp. Dev. 19:4, pp. 712–728 (2012)
16. Hwang, M.: Disentangling the Effect of Top Management Support and Training on Systems Implementation Success: A Meta-Analysis. Commun. AIS. 35:2, pp. 19–37 (2014)
17. Harris, M., and Weistroffer, H.: A New Look at the Relationship between User Involvement in Systems Development and System Success. Commun. Assoc. Inf. Syst. 24, pp. 739–756 (2009)
18. Hung, S.-Y., Chen, C., Wang, K.-H.: Critical success factors for the implementation of integrated healthcare information systems projects: An organizational fit perspective. Commun. Assoc. Inf. Syst. 34:1, pp. 775–796 (2014)
19. Venkatesh, V., Bala, H., Sambamurthy, V.: Implementation of an Information and Communication Technology in a Developing Country: A Multimethod Longitudinal Study in a Bank in India. Inf. Syst. Res. 27:3, pp. 558–579 (2016)
20. Mahmud , I., Ramayah, T. , Kurnia, S.: To use or not to use, Inf. Syst. 69, pp. 164–179 (2017)
21. Ahuja, S., Chan, Y.E.: The Enabling Role of IT in Frugal Innovation. In: Proceedings of the Thirty Fifth International Conference on Information Systems (ICIS). pp. 1–20 (2014)

22. Pries-Heje, L., Dittrich, Y.: ERP Implementation as Design: Looking at participatory design for means to facilitate knowledge integration. Scand. J. of Inf. Syst. 21:2, pp. 27–58 (2009)
23. Srivastava, S.C., Shainesh, G.: Bridging the service divide through digitally enabled service innovations: Evidence from indian healthcare service providers. MIS Q. 39:1, pp. 245–267 (2015)
24. Lusch, R.F., Nambisan, S.: Service Innovation: A Service-Dominant Logic Perspective. MIS Q. 39:1, pp. 155–175 (2015)
25. Vargo, S.L., Maglio, P., Akaka, M.: On value and value co-creation: A service systems and service logic perspective. Eur. Manag. J. 26:3, pp. 145–152 (2008)
26. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. J. Mark. 68:1, pp. 1–17 (2004)
27. Matzner, M., Büttgen, M., Demirkan, H., Spohrer, J., Alter, S., Fritzsche, A., Ng, I.C.L., Jonas, J.M., Martinez, V., Möslein, K.M., Neely, A.: Digital Transformation in Service Management. J. Serv. Manag. Res. 2:2, pp. 3–21 (2018)
28. Ko,D.-G., Laurie J.K., William R.K.: Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations. MIS Q. 29:1, pp. 59–85 (2005)
29. Agarwal, N., Brem, A.: Frugal innovation-past, present, and future. IEEE Eng. Manag. Rev., 45:3, pp. 37–41 (2017)
30. Bhatti, Y.A., & Ventresca, M.: The Emerging Market for Frugal Innovation: Fad, Fashion, or Fit? SSRN Electronic Journal. (2012)
31. Heeks, R., Foster, C., Nugroho, Y.: New models of inclusive innovation for development. Innov. Dev. 4:2, pp. 175–185 (2014)
32. Agarwal, N., & Brem, A.: Frugal and reverse innovation - Literature overview and case study insights from a German MNC in India and China. In: Proceedings of the 18th International ICE Conference on Engineering, Technology and Innovation, pp. 1–11 (2012)
33. Baker, T., Nelson, R.E.: Creating Something from Nothing: Resource Construction through Entrepreneurial Bricolage. Adm. Sci. Q. 50:3, pp. 329–366 (2005)
34. Radjou, N., & Prabhu, J.: Frugal innovation: how to do more with less. London: Economist (2016)
35. Zeschky, M., Widenmayer, B., Gassmann, O.: From Cost to Frugal and Reverse Innovation: Mapping the Field and Implications for Global Competitiveness. Res. Technol. Manag., 57:4, pp. 20–27 (2014)
36. Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V., Bragge, J.: The Design Science Research Process: A Model for Producing and Presenting Information Systems Research. In: Proceedings of Design Research in Information Systems and Technology (DESRIST), pp. 83–106 (2006)
37. Webster, J, Watson, R.: Analyzing the past to prepare for the future: Writing a literature review. MIS Q. 26:2, pp. 13–23 (2002)
38. Chauhan, S., Jaiswal, M.: Determinants of acceptance of ERP software training in business schools: Empirical investigation using UTAUT model. Int. J. Manag. Edu. 14:3, pp. 248–262 (2016)
39. Hwang, M., Cruthirds, K.: Impact of an ERP simulation game on online learning. The Int. J. Manag. Edu. 15:1, pp. 60–66 (2017)
40. Prabhu, J., Jain, S.: Innovation and entrepreneurship in India: Understanding jugaad. Asia Pac. J. Manag. 32, pp. 843–868 (2015)
41. Sirmon, D.G., Hitt, M.A., Ireland, R.D., Gilbert, B.A.: Resource orchestration to create competitive advantage: Breadth, depth, and life cycle effects. J. Manag. 37:5, pp. 1390–1412 (2011)

42. Watson, R.T., Kunene, K.N., & Islam, M.S.: Frugal information systems (IS). Inf. Technol. Dev., 19:2, pp. 176–187 (2013)
43. IfM (Institut für Mittelstandsforschung): Mittelstand im Überblick, http://www.ifm-bonn.org/statistiken/mittelstand-im-ueberblick/#accordion=0&tab=0 (Accessed: 15.11.2018)
44. Hausman, A.: Innovativeness among small businesses: Theory and propositions for future research. Ind. Mark. Manag. 34:8 , pp. 773–782 (2005)
45. Offermann, P., Levina, O., Schönherr, M., & Bub, U.: Outline of a design science research process. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST), pp. 1–11 (2009)
46. Etikan, I., Musa, S. & Alkassim, R.: Comparison of Convenience Sampling and Purposive Sampling. Am. J. of Theoretical and Applied Statistics. Vol. 5, No. 1, 2016, pp. 1–4 (2016)

How to Structure a Company-wide Adoption of Big Data Analytics

Olga Bürger[1,2]

[1] University of Augsburg, FIM Research Center, Augsburg, Germany
olga.buerger@fim-rc.de
[2] Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Augsburg, Germany

Abstract. Driven by increasing amounts of data and by emerging technologies to store and analyze them, companies adopt Big Data Analytics (BDA) to improve their innovativeness and decision-making. However, adopting BDA across the company in the sense of an insight-driven organization (IDO) is challenging, since it influences the entire company and requires an organizational change. Despite mature knowledge, approaches that provide concrete methods for structuring the company-wide adoption of BDA to fully exploit the benefits of BDA and to reduce the risk of its failure are still missing. Following action design research, we developed and evaluated a method for structuring the company-wide adoption of BDA in a concerted research effort at a German bank. Based on knowledge of BDA and the roadmapping approach, the method structures the adoption along the BDA capabilities. We illustrate how companies can define a target state, identify gaps, and derive a BDA roadmap.

Keywords: Big Data Analytics, Roadmapping, Action Design Research.

1 Introduction

Inspired by big players, such as Google and Amazon, companies increasingly adopt Big Data Analytics (BDA) as an approach to utilize big data and advanced analytics for delivering value, improving efficiency, and establishing competitive advantage [1], [2]. The rapid growth of data generated by social media like Facebook and Twitter, as well as emerging information technologies (IT) like the Internet of Things, advance this trend [2], [3]. Additionally, the market increasingly offers more mature and powerful tools to source, store, and analyze big data, which lay the foundation for adopting BDA. Considering its technological capabilities and the associated high expectations, it is not surprising that BDA is – meanwhile – considered as a game changer, due to its operational and strategic potential [1]. Thereby, adopting BDA across the whole company instead of only using it within individual projects is considered as more beneficial since companies that use BDA as a competitive differentiator, – defined as insight-driven organizations (IDOs), – tend to perform better in terms of financial and operational results [4], [5]. They are champions in implementing BDA and improving the speed and quality of action through data-driven decisions [4], [5].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

However, a company-wide adoption of BDA is challenging as it requires a long-term evolution [3], involves different stakeholder groups, impacts various levels of the enterprise architecture, and needs high investment amounts [6]. Due to this complexity, a structuring approach is important to coordinate the individual measures, taking into account the dependencies in terms of content and time. Prior research has already revealed factors that may be relevant for a structured adoption of BDA. For example, [7] address the need for a clear vision of what companies want to achieve and a roadmap to reach the target. [4] concretize that companies should define the business challenges, identify the organizational changes needed, and derive a roadmap. However, they do not show how companies can apply this procedure. Furthermore, prior studies recommend the development of BDA maturity or capability models that allow companies to assess their current state regarding the required capabilities [8]. However, they fail to illustrate how these models can be used for a coordinated company-wide adoption of BDA. Finally, [5], [9] advise companies to start with seed or lighthouse projects for a few use cases to gain initial experience with BDA, encourage collaboration, and create awareness. Thereby, they do not consider the long-term changes. Thus, despite addressing important issues, approaches that provide concrete methods for structuring the company-wide adoption of BDA are still missing. In order to contribute to this research gap, we study the following research question: *How can developing a roadmap assist in structuring the company-wide adoption of BDA?*

In order to answer this question, we adopt action design research (ADR) and develop a method that aims to assist companies in structuring a company-wide adoption of BDA. In line with ADR, we develop and evaluate our method in a concerted research effort at a German bank [10]. The paper is organized as follows: Section 2 contains a brief overview of the related research. After introducing our methodology in Section 3, we describe our method's design in Section 4 and evaluate it in Section 5. We conclude by discussing implications, limitations, and directions for future research in Section 6.

2 Background and Related Work

Since our research is motivated by a concrete problem in practice, we have first discussed their needs with the end-users to achieve an in-depth understanding of the problem, and then researched the literature for appropriate methods to solve it. Therefore, this section provides a brief overview of the work related to BDA as the main content of the project and to roadmapping as one possible concept to approach the solution of the problem. As already recommended by prior research (e.g. [4], [7]), we focus on roadmapping to structure the adoption of BDA since it allows to define a target state, to identify gaps, and to derive and prioritize measures to reach the target state.

2.1 Big Data Analytics (BDA)

Prior research states that developing appropriate BDA capabilities can help companies to successfully adopt BDA in order to become an IDO. Thereby, studies define different BDA capabilities. For example, [11] identify culture, data management, and skills as the main dimensions of a BDA capability, whereas [1] define BDA infrastructure, personnel, and management capability as the key components. [8] identify thirty four generic capabilities, which they assign to eight capability fields (e.g. customer relationship management, strategy development, and transformation competence). Thereby, they state that the relevance of the capabilities might vary, depending on the scenario. Despite differences about the identified components and the level of granularity, all studies have a multidimensional perspective and address the need to develop BDA capabilities to successfully adopt BDA. Further, [4] provide an approach that could be the first step toward a BDA maturity model. They consider people, processes, and tools as the necessary building blocks for BDA and define three levels of BDA capability: aspirational, experienced, and transformed. Whereas aspirational organizations focus on process efficiency or process automation to cut costs, experienced organizations aim at optimizing their organizations by developing new ways to use BDA. Transformed organizations (i.e. IDOs) focus on using BDA as a competitive differentiator to expand their market position [4]. During the transformation, BDA expands from use in only selected business units toward organization-wide adoption [4], [9]. Further studies highlight the need for managing transformation effectively [5] and structuring it by defining a clear vision, identifying required changes, and deriving a roadmap to achieve the target state [4], [7]. Since the company-wide adoption of BDA might require a long-term evolution, companies can start with lighthouse projects for selected business units to gain initial experience with BDA [5], [9] and provide initial results by using prototyping.

2.2 Roadmapping

The roadmapping approach is a widely used management concept for supporting strategy and innovation [12]. It has been widely adopted at various levels of granularity from product to industry sector and also across various industries [6], [12]. Thereby, roadmaps can be used to communicate visions, to explore the development of the business and its components, to coordinate activities and resources, and to monitor progress [12], [13]. They also enable the alignment of different functions and perspectives within an organization, particularly business and technology [12]. The roadmaps are also very flexible and scalable, and can be customized to suit different strategic and innovation contexts [14]. The most general approach delivers a framework that addresses three key questions: 1) Where do we want to go? 2) Where are we now?, and 3) How can we get there? [12]. The first key question refers to the definition required for a target vision, the second one aims at covering the gaps between the status quo and the target vision, and the third one includes identifying, as well as structuring, the measures for achieving the target vision. The roadmap architecture consists of two dimensions: 1) timeframes, as well as 2) layers and sub-layers [12]. Timeframes are usually the horizontal axis and include short-, medium-, and long-term perspectives, as well as the past and vision. The layers and sub-layers are usually the vertical axis and show different levels of a hierarchical taxonomy. Since roadmaps can be used at various levels of granularity, the roadmap architecture should be customized to suit the aims and scope of the contemplated effort [12].

3 Research Method

In order to answer our research question, we relied on Action Design Research (ADR) to build, intervene, and evaluate our method. ADR involves the construction of an artefact, its intervention in the organization, and its evaluation by means of a concerted effort [10]. The developed artefact reflects, therefore, not only the theoretical precursors and the researchers' intent, but also the users' influence in organizational contexts [10]. Using ADR enabled us to design and fine-tune our method such that we could provide both academic insights and practical value. We implemented the four ADR stages. According to the first stage, – problem formulation, – we studied the research gap in the existing knowledge and outlined our research question in the introduction [10]. The second stage involves building, intervention, and evaluation activities. During this stage, we designed and evaluated our method at a German bank. In the third stage, – reflection and learning, – we continuously reflected on our method's design and analyzed the intervention results in context of our method's goals by integrating the feedback received from practitioners and end-users. In the fourth and last ADR stage that aims to formalize the learning gained throughout the project, we identified general insights about activities and techniques (cf. introducing our method below). In order to ensure that our method includes the relevant attributes and elements needed to design a new method, we further relied on the mandatory method components provided by [15] as shown in Table 1.

Table 1. Mandatory Method Components [15]

	Name	*Description*
Attributes	(A.1) Goal orientation	Methods must strive for achieving specific goals
	(A.2) Systematic approach	Methods must include a systematic procedure model
	(A.3) Principles orientation	Methods must follow general design guidelines and strategies
	(A.4) Repeatability	Methods must be repeatable in different contexts
Elements	(E.1) Activity	Task that creates a distinct (intermediate) output
	(E.2) Technique	Detailed instruction that supports the execution of an activity
	(E.3) Tool	Tool (e.g. method) that supports the application of a technique
	(E.4) Role	Actor that executes or is involved in the execution of an activity
	(E.5) Defined output	Defined outcome per activity (e.g. artefact, documents)

4 The Method Design

4.1 Design Principles

In line with ADR, we derive design principles for our method [16] from the existing theory and knowledge gained during the project [10]. As detailed above, a company-wide adoption of BDA is challenging as it affects various levels of the enterprise architecture and involves different stakeholder groups [6]. Thus, the company-wide adoption of BDA needs a clear vision of the target state, as well as a concept to capture the status quo, and to identify the changes needed to reach the target [4], [5], [7]. Thereby, a BDA capability model can provide guidance on which capabilities an organization should develop to become an IDO [4], [8]. This leads us to define the following design principle: *(DP.1) The method should allow for a precisely defined target state to be achieved by the adoption of BDA and identification of measures to close the gaps between the status quo and the target state. It should further take into account BDA capabilities needed as well as various levels of the enterprise architecture and stakeholder groups.* Besides, organizations also need guidance on how to proceed to reach the target state [4], [7] by prioritizing and structuring the identified measures [6] to coordinate the initiatives with regard to limited resources and predecessor-successor dependencies. As the company-wide adoption of BDA requires a long-term evolution [3] and high investments [6], definition of milestones might help to reevaluate and terminate the transformation project, if necessary. We therefore define the following design principle: *(DP.2) The method should allow for prioritizing and structuring the implementation measures according to the BDA capability developed by them and to the time of their implementation. It should further enable defining the milestones for reevaluation.*

4.2 Method Procedure Model

In keeping with [15], our method consists of activities (E.1), each of which includes techniques (E.2), tools (E.3), roles (E.4), and output (E.5) as summarized in Table 2. Our method comprises three activities: defining the target state, identifying and prioritizing the gaps, and deriving a BDA roadmap. Although tools can be defined as IT tools only, we use a broader definition of [15], [16] and focus on tools that support the application of one or more techniques.

Activity 1: Defining the target state

Technique: According to [12], activity 1's purpose is to address the question "Where do we want to go?" and to define the target state to be achieved by the adoption of BDA. First, the method user needs to define a target vision to have a common understanding of the target state. For companies that aim at using BDA as a competitive differentiator, becoming an IDO can be an appropriate target vision. In order to avoid a target vision being almost unattainable or only achievable with a great deal of effort, the method user can derive a second target vision as an intermediate step positioned between the status quo and an IDO. Furthermore, the intermediate step might allow a reevaluation of the targets and even a termination of the project, if necessary. Based on the defined target vision, the method user should define requirements that need to be fulfilled at the target state and group them according to appropriate dimensions to conceptualize the target state. Later on, these dimensions will be visualized as layers in the BDA roadmap. Since development of BDA capabilities can be an appropriate way to become an IDO, we recommend that selected BDA capabilities be used as roadmap layers. After conceptualizing the target state, the method user should operationalize it by breaking down the requirements into fields of action and group them into roadmap layers or sub-layers. The number of sub-layers should meet the appropriate degree of granularity. While too many sub-layers lead to a very detailed and overloaded roadmap, too little sub-layers would make it difficult to derive effective measures to close the gaps [12]. We recommend deriving a maximum of 5 – 8 sub-layers for any layer [12].

Tool: We recommend that all the activities of our method should be based on the roadmapping as a structuring approach [12] and techniques such as brainstorming and moderated group discussions [15] to generate and evaluate ideas. In order to derive layers and sub-layers of the roadmap, the method user can base on BDA capability models (e.g. [1], [8], [11]).

Roles: In order to carry out all activities of our method, we recommend that a project team, which can consist of internal and / or external experts in BDA and developing roadmaps, be assigned. The project team prepares and moderates discussions, interviews, and workshops. They also consolidate and analyze the input, and provide outputs. Since management support is a well-known success factor for projects with high transformation potential like the company-wide adoption of BDA [4], [6], activity 1 involves (senior) managers who are familiar with the organization's strategy.

Output: Activity 1's output is the target state(s) as well as fields of action grouped into layers and sub-layers.

Table 2. Overview of Method's Activities and Elements

Activity	*Technique*	*Tool*	*Role*	*Output*
Activity 1: Defining the target state	Define the target state based on selected dimensions	Roadmapping, BDA capability model, discussion	(Senior) Managers, project team	Target state, layers, sub-layers, fields of action
Activity 2: Identifying and prioritizing the gaps	Capture the status quo, identify and prioritize the gaps	Semi-structured interviews, fulfillment-importance matrix	Stakeholder, (senior) managers, project team	Fulfillment-importance matrix with prioritized gaps
Activity 3: Deriving a BDA roadmap	Derive and structure measures to close the gaps	Roadmapping	(Senior) Managers, project team	BDA roadmap with structured measures

Activity 2: Identifying and prioritizing the gaps

Technique: Consistent with [12], activity 2 aims at addressing the question "Where are we now?" as well as identifying and prioritizing the gaps between the status quo and the target state. In the first step, the method user needs to identify the experts who can give input on the derived fields of action (cf. *Roles* below). In the next step, they need to collect quantitative and qualitative data for further analysis to identify and prioritize the gaps by, for example, using the tools described below.

Tool: We recommend using semi-structured interviews [17]. The method user can include selected follow-up questions that match with the interviewees' areas of expertise to gain more insights. They can also include overarching questions to bring out the interviewees' expectations concerning their perceived challenges and opportunities regarding the company-wide adoption of BDA. In order to assess the status quo of fields of action in terms of their relevance and degree of fulfilment, we recommend using five-point Likert scales with 1 = irrelevant / not fulfilled at all and 5 = highly relevant / completely fulfilled. For identifying and prioritizing the gaps, the method user can adapt the fulfillment-importance matrix, which slightly resembles a mirrored version of Gartner's Magic Quadrant [18]. Therefore, they need to assign the fields of action according to their assessment of the matrix's four quadrants: "Invest!", "Manage Excellence!", "Reprioritize! or Disinvest!", and "Ignore!". The fields of action in the "Invest!" quadrant are the most important gaps and need to be closed by moving the associated requirements to the "Manage Excellence!" quadrant.

Roles: We recommend that the project team prepares, conducts, and analyzes the interviews. In order to gain sufficient information for a comprehensive report on the status quo, experts from different management layers and stakeholder groups should be interviewed (e.g. IT, finance, risk management, and sales departments). If necessary, the project team should consult the (senior) managers to identify the appropriate experts. Internal experts from projects with a similar focus (e.g. data quality projects) as well as external experts (e.g. consultants) that accompany these projects can also be interviewed. Each interview should be conducted by at least two project team members to avoid subjective interpretations of the answers.

Output: Activity's 2 output is a fulfillment-importance matrix with prioritized gaps.

Activity 3: Deriving a BDA roadmap

Technique: Following [12], activity 3 aims at addressing the key question "How can we get there?", as well as identifying the measures to close the gaps and structuring them in a roadmap. Method user should derive the measures and assign them according to the roadmap's sub-layers and layers. In terms of timeframe, the measures need to be structured according to their short-, medium-, and long-term perspective. Since the company-wide adoption of BDA requires a long-term evolution [3], a BDA roadmap might have a timeframe that spans over several years. Thus, the method user should consider intertemporal and scheduling interactions between the measures [6].

Tool: For identifying appropriate measures to close the gaps, method user can rely on knowledge about IDO, BDA, and BDA capabilities as well as brainstorming and discussions within the project team. For the latter, method user should be aware of limitations of relying on existing personal knowledge of the involved practitioners. A close collaboration between researchers and practitioners can help to reduce this bias if an intensive and reflective discussion process is ensured to combine the perspectives and insights from researchers and practitioners. For structuring the measures, we recommend deriving a roadmap as an established planning tool [12]. Thereby, the layers and sub-layers are based on the BDA capabilities derived in activity 1. The timeframe includes short-, medium-, and long-term perspectives.

Roles: The project team derives the measures to close the gaps from the literature, structures them in a roadmap, and evaluates the results with the (senior) managers.

Output: The output is a BDA roadmap with structured measures.

5 Evaluation

5.1 Case Study Setting

We conducted our study in the strategic department of one of the leading universal banks in Germany. Since the banking industry is exposed to increasing innovation pressure through changing client behavior [19] and new market players like FinTechs [20], financial service providers need to innovate their current value delivery and interactions with clients [21] through providing data-driven services, for example. As a financial service provider, the case-study bank particularly had a large volume of client data such as details on repayment behavior and outstanding loans or credit lines. However, the bank failed to systematically get value from its data and it also failed to provide data-driven services to its clients. Thus, the bank aimed at adopting BDA across the whole company to develop it towards an IDO and to capture BDA potentials like strengthening innovative power and improvement of decision-making. However, the bank faced various challenges for this project. For example, a lack of end-to-end processes, a lot of manual work involved in the processes and thinking in silos led to frequent breaks in information flows. In addition, the bank had a partially outdated IT architecture and outsourced IT. Finally, the bank put a lot of effort into regulatory-driven projects and thus had rather limited human and financial resources for innovation projects. This led to a lack of awareness for innovation projects in general and for a company-wide adoption of BDA in particular.

The objective of the case study was therefore to develop a method for structuring the company-wide adoption of BDA to shift the company toward an IDO. Thereby, the head of the strategy department and the CIO, who recognized the market relevance of BDA, were looking for a method that would allow them to show the potentials of BDA to create awareness and to include the entire organization as well. With regard to the challenges highlighted above, the new method should consider various perspectives (e.g. people and processes) to enable including different stakeholder groups and levels of the enterprise architecture. Since the board of directors would have to approve the new initiative, the new method should also allow to assess what exactly they want to achieve with the company-wide adoption of BDA, what the status quo looks like and how they aim to achieve the target state. Due to the high complexity of the project, the new method should provide guidance in covering which adoption measures are necessary, how long the adoption will last, whether there are dependencies between the measures and which resources will be necessary for their implementation.

The project team consisted of five academic members from the authors' institutions (two research fellows and three professors) with expertise in developing digital roadmaps and in the financial services industry. In addition, the project team included four members of a consulting company (two consultants and two senior consultants) with BDA expertise and experience in regulatory-driven projects conducted at the case-study bank. The head of the strategy department and his assistant were also part of the team. The external project team's role was to prepare and conduct workshops, as well as interviews, and to develop the BDA roadmap. The project lasted three months in total. The external team members mostly worked in the back office and were on site for workshops, interviews, and other meetings. The external project team predominantly worked three to four days a week on the project and spent the rest of the time synchronizing with colleagues who were working in similar areas.

5.2 Method Application

Activity 1: Defining the target state

In order to create a common understanding of the target vision, we first discussed the meaning of an IDO with the head of the strategy department and the CIO. Thereby, we defined an IDO as follows: An IDO anchors data-based decision-making processes throughout the company and classifies big data as a core capability with the aim of making value-creating insights available at the right place and at the right time. According to the feedback of the bank end-users, we derived a second target vision as an intermediate step on the way from the status quo to an IDO to take into account the current challenges of the bank (e.g. a lack of resources and awareness). Since a company-wide adoption of BDA requires high investments, this step was also defined as a milestone for reevaluating the targets and the project to adjust the resource allocation or to determine the project, if necessary. In order to better express the focus of both target visions, we named them "Lab" and "Factory". Following the recommendations of prior research (e.g. [4], [5], [9]), the target vision "Lab" aimed at using BDA in selected business units – mostly in the form of lighthouse projects, – which should serve as enabler of strategic corporate goals. The target vision "Factory" focused on a group-wide use of BDA as a competitive advantage and unique selling proposition in an IDO context. In the next step, we conceptualized the target states by deriving the roadmap layers based on five BDA capabilities of strategy, people, process, data, and technology. Since a company-wide adoption of BDA influences different levels of the enterprise architecture, the BDA capabilities should be oriented at these levels. With this in mind, we first drew up a list of possible BDA capabilities based on literature research and selected the most important ones for the bank in a workshop with the head of the strategy department. For example, developing strategy and people capabilities might help to increase company-wide awareness for BDA, whereas process capabilities would allow accelerating innovation and decision-making processes. Since BDA requires an excellent handling of data and advanced technologies to collect, store and analyze it, the bank finally needed to develop data and technology capabilities. In close collaboration with the head of the strategy department, we assigned the defined target visions to the layer strategy. We then derived requirements to be fulfilled at the target states and grouped them into the remaining layers as summarized in Table 3. Thereafter, we specified the requirements by breaking them down into the fields of action and grouping them into roadmap sub-layers. In sum, we derived twenty four fields of action (e.g. specialists, research environment, idea generation, and agility) and seven sub-layers (team structure, broad knowledge, innovation process, project management, data quality, data access and trust, and toolkit). Due to the challenges described above, several fields of action aimed at improving innovation processes (in particular increasing their speed) and promoting interdisciplinary collaboration to counter silo thinking.

Table 3. Target States at the Case-Study Bank with Selected Requirements

Layers	*"Lab"*	*"Factory"*
Strategy	Use of BDA in selected business units in the form of lighthouse projects which serve as enabler of strategic corporate goals	Group-wide use of BDA as a competitive advantage and unique selling proposition in the sense of an IDO
People	Specialists in selected functions	Specialists in all functions
Process	Focus on lighthouse projects to create awareness	Integration into daily processes
Data	Data standardization within the lighthouse projects	Data lifecycle management
Technology	Focus on visualization software	Use of advanced BDA tools

Activity 2: Identifying and prioritizing the gaps

We conducted eleven semi-structured interviews with heads and members of different departments (e.g. finance and IT), as well as internal and external experts involved in regulatory-driven projects (e.g. AnaCredit and BCBS#239). The interviews followed the fields of action defined in activity 1. In close collaboration with the head of the strategy department, we excluded the strategy capability from the questionnaire, because it was defined through the target visions. Each interview was conducted by at least two external project team members and lasted roughly one hour. The interviewed experts answered the questions for each field of action and assessed the relevance, as well as degree of fulfilment, on a scale of 1 to 5 with 1 = irrelevant / not fulfilled at all and 5 = highly relevant / completely fulfilled. We also included follow-up and overarching questions to gather qualitative data for further analysis.

In the next step, we aggregated the quantitative results for each field of action and assigned them to quadrants of a fulfillment-importance matrix as shown in Figure 1. According the five-point Likert scale, we defined the quadrants of the fulfillment-importance matrix by interpreting fulfillment values less than or equal to 3 and importance values less than 3 as low. We treated all fields of action located in the "Invest!" quadrant as gaps with a high priority and the fields of action assigned to the "Manage Excellence!" quadrant as gaps with medium to low priority. Based on our analysis, there were no fields of action assigned to the "Reprioritize! or Disinvest!" and "Ignore!" quadrants. This is reasonable, because evaluating the fields of action with practitioners and end-users in activity 1 already ensured that the most relevant fields of action were identified, also considering the organizational idiosyncrasies. We further evaluated the matrix by analyzing the qualitative insights from the follow-up, as well as overarching questions, and discussing our results with the head of the strategy department. Our first result was that the bank did well in a few central topics of the company-wide adoption of BDA, since many fields of action were located in the “Manage Excellence!” quadrant. For example, the bank made an effort to establish a data quality awareness within regulatory-driven projects. Most interestingly, the fields of action located in the “Invest!” quadrant were distributed almost equally across the four BDA capabilities. Therefore, deriving a BDA roadmap as a purely IT-driven effort would have neglected a substantial share of the relevant gaps. The gap analysis revealed that the first step was laying the foundations for an IDO by, for example, establishing a team of specialists, introducing an explicit research environment, and adopting basic technologies. On this basis, the bank could then begin with the more culture-oriented shift toward an IDO by focusing on generating innovative ideas, agility, and speed when it comes to the implementation of ideas.

Figure 1. Fulfillment-Importance Matrix at the Case-Study Bank

Activity 3: Deriving a BDA roadmap

We derived measures to close the gaps based on the literature review, as well as discussions with the head of the strategy department and the CIO. In the next step, we structured the measures in the BDA roadmap in terms of sub-layers and timeframe-dimensions. We also delivered a comprehensive documentation with a detailed description of each measure. For anonymization reasons, Figure 2 shows only a high-level BDA roadmap with selected measures structured into five layers (i.e. strategy, people, process, data, and technology) and three timeframe-dimensions (i.e. short-term phase 1, medium-term phase 2, and long-term phase 3). The target vision "Lab" should be reached at the end of phase 2, and the target vision "Factory" at the end of phase 3. Thereby, the planning reliability and granularity of measures are greatly reduced in phase 3 due to its long-term focus. We also included phase 0, which indicates the project start in the current year, as well as two evaluation loops at the end of phase 1 and phase 2 as a reevaluation or termination option.

In order to reach the target "Lab", within phase 1 and phase 2, we structured the measures aimed at creating BDA awareness and initiating a data-driven culture via lighthouse projects. These measures include, for example, recruiting internal and external specialists, as well as providing a research environment and technologies to carry out initial lighthouse projects. Further, measures like conducting lighthouse projects, providing the first prototypes through agile methods, and design thinking should foster idea generation, organizational agility, and speedy ideas implementation. The measures in phase 3 aimed at closing the gaps to reach the target vision "Factory". Measures like organization-wide training programs in BDA or agile (innovation) project management approaches should ensure innovativeness and speed. Organization-wide data quality measures need to enable a high-quality data as basis for BDA. Further, measures like the implementation of a data lake architecture and a central sandbox should provide a flexible and scalable technological base that e.g. allows for adopting various BDA technologies. In close collaboration with the head of the strategy department, we also included strategy-related measures in our roadmap (e.g. change management, as well as strategic alignment between BDA measures and ongoing IT and regulatory-driven projects).

	Start Phase 0	**Lab** Phase 1 (short-term)	Phase 2 (medium-term)	**Factory** Phase 3 (long-term)
Strategy	• Change management: Promoting agility & innovation, embedding BDA in corporate culture, developing a CoE for BDA • Strategic alignment: Closely interlinking BDA and IT roadmaps as well as regulatory projects			
People	Set-up of the BDA transformation project • Kickoff • Detailed planning • Concretization of the roadmap • Preparation and start of initial measures	• Internal and external specialists • Trust through lighthouse projects	• Individually training programs • BDA goal for middle management	• Training programs for specialists • Company-wide education programs
Process		• Agile project methods • Simplified administrative processes	• Innovation management for BDA • Explicit BDA process	
Data		• Mirrored database • Program for data quality	• Data integration within the pilot areas • DQM in the pilot areas	• Enterprise data framework • Company-wide DQM
Technology		• Software for the lighthouse projects • Prototypes for lighthouse projects	• Software deployment process • Lighthouse projects	• Cloud for high-performance analysis • Systematic adoption of new technologies

Completion & evaluation of first 2 lighthouse projects

Evaluation achievement target vision Lab & decision on expansion to Factory

Figure 2. High-Level BDA Roadmap at the Case-Study Bank

5.3 Method Evaluation

Regarding the evaluation of design artefact, we can state that the new method that we co-developed and applied at a German bank provides an initial proof-of-value, since it fulfills the requirements of the bank outlined in Section 5.1. In particular, the new method enabled the bank to structure the company-wide adoption of BDA by deriving a roadmap that considers various perspectives (i.e. strategy, people, process, data, and technology) and prioritizes measures to close the identified gaps. Furthermore, the end-users evaluated the new method as understandable and practicable. Moreover, the board of directors accepted it and the bank has already started initiatives to implement first projects proposed in the roadmap. From a more abstractive point of view, our method fulfills the content and domain-specific requirements defined by the two design principles (DP1 and DP 2) in Section 4.1. According to DP1, our method allows defining one or more target states and their operationalization by considering selected BDA capabilities as shown in activity 1. In activities 2 and 3, it enables identifying gaps and deriving measures to close these gaps by involving different stakeholder groups (e.g. through interviews). Finally, the new method considers various levels of the enterprise architecture when defining the layers and sub-layers of the BDA roadmap. According to DP 2, our method allows prioritizing and structuring the implementation measures in a BDA roadmap as illustrated in activity 3. It also enables considering dependencies between individual measures and defining the milestones for the project reevaluation. Finally, our method meets general requirements for a new method because it contains the mandatory method components summarized in Table 1. In terms of goal orientation, our method aims at structuring the company-wide adoption of BDA. As for principles orientation, our method is geared toward two design principles derived from the literature and fine-tuned with practitioners and end-users within an ongoing method evaluation by incorporating requirements outlined in Section 5.1. Repeatability and systematic approach are achieved by describing the method procedure model in detail and demonstrating its applicability at the case-study company.

Regarding the evaluation of design process, our method design follows the seven ADR principles. Within the first stage, we followed the ADR principle of practice-inspired research by illustrating that practice pays a lot of attention to BDA adoption. As for the ADR principle of theory-ingrained artefacts, our method bases on the existing knowledge related to BDA and roadmapping. During the second stage, we followed the ADR principles of reciprocal shaping and mutually influential roles, as well as authentic and concurrent evaluation by co-developing and evaluating our method in an iterative manner with the practitioners and the bank end-users. Through a continuous reflection on our method's design within the third stage, our method does not only reflect the preliminary design, but also the organizational shaping, as well as the practitioners' and end-users' feedback, thereby meeting the ADR principle of guided emergence. For example, activity 1 initially included defining one target state. After an evaluation with the bank end-users, we added the opportunity of defining more target states including a short explanation when it might be useful. In the fourth stage, we followed the ADR principle of generalized outcomes by providing general insights about activities for structuring a company-wide adoption of BDA.

6 Conclusion

In this study, we investigated how organizations can structure the company-wide adoption of BDA. Using ADR, we developed a new method for structuring the company-wide adoption of BDA by deriving a BDA roadmap. Based on knowledge of roadmapping and BDA, our method includes three activities: 1) defining the target state, 2) identifying and prioritizing the gaps, and 3) deriving a BDA roadmap. Consistent with ADR, we developed and evaluated our method in a concerted research effort at a German bank.

Our work contributes to both research and practice. From an academic perspective, we enrich the body of knowledge related to BDA by linking the concept of BDA capabilities with the roadmapping approach when developing a new method for structuring the company-wide adoption of BDA. In particular, we show how companies can develop a BDA roadmap by considering BDA capabilities. Furthermore, we extend prior research on BDA capabilities by applying the concept of BDA capabilities to a concrete use case and illustrating how this concept can help to structure the company-wide adoption. Thus, our work can serve as a starting point for developing BDA maturity models and investigating their application in practice. Practitioners can use our method as a guideline for structuring the company-wide adoption of BDA. They can customize our method by extending our dimensions based on the BDA capabilities or by using other dimensions. Moreover, our research might help to develop company-individual methods for structuring other complex efforts like innovation and business transformation projects.

Our research has limitations that can serve as further starting points for future studies. First, we derived a customized BDA roadmap and noticed a lack of holistic BDA capability models that can be addressed by further research. Further, our method focuses on deriving a BDA roadmap as a planning tool and neglects the implementation phase. Research based on successfully carried out but also failed BDA adoption projects could be helpful for ex post analyses of success factors and development of key performance indicators to manage the adoption. Developed and evaluated at a German bank, our method is to a certain extent company-specific. Nevertheless, many aspects of our method can be generalized. As in our case, organizations should ensure a multidimensional view of the BDA adoption. Our experience also corroborates the importance of a close collaboration between strategy department, IT department, and business units, as well as the roadmap alignment to ongoing IT and regulatory-driven projects. Conducting further case studies might provide further valuable insights and outline possible differences along industries or the type of adoption projects. Despite its limitations, our research postulates a method for structuring the company-wide adoption of BDA and serves as a basis for further research aimed at closing the outlined research gap.

References

1. Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.F., Dubey, R., Childe, S.J.: Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research. 70, 356-365 (2017)
2. Müller, O., Junglas, I., vom Brocke, J., Debortoli, S.: Utilizing big data analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems. 25(4), 289-302 (2016)
3. Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., Zhao, J.L.: Transformational Issues of Big Data and Analytics in Networked Business. MIS quarterly. 40(4), 807-818 (2016)

4. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big Data, Analytics and the Path from Insights to Value. Sloan Management Review. 52(2), 20-31 (2011)
5. McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D. J., Barton, D.: Big data: the management revolution. Harvard business review. 90(10), 60-68 (2012)
6. Röglinger, M., Häckel, B., Bolsinger, M., Walter, M.: How to Structure Business Transformation Projects: The Case of Infineon's Finance IT Roadmap. Journal of Information Technology Theory and Application. 17(2), 5-21 (2016)
7. Kiron, D.: Organizational Alignment is Key to Big Data Success. MIT Sloan Management Review. 54 (3), 1-6 (2013)
8. Dremel, C., Overhage, S., Schlauderer, S., Wulf, J.: Towards a Capability Model for Big Data Analytics. In: Proceedings of the 13th International Conference on Wirtschaftsinformatik (WI), pp. 1141-1155. St. Gallen (2017)
9. Gust, G., Flath, C.M., Brandt, T., Ströhle, P., Neumann, D.: How a Traditional Company Seeded New Analytics Capabilities. MIS Quarterly Executive. 16(3), 215-230 (2017)
10. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS quarterly. 35(1), 37-56 (2011)
11. Kiron, D., Prentice, P.K., Ferguson, R.B.: The analytics mandate. MIT Sloan management review, 55(4), 1-25 (2014)
12. Phaal, R., Muller, G.: An architectural framework for roadmapping: Towards visual strategy. Technological Forecasting and Social Change. 76(1), 39-49 (2009).
13. Kostoff, R.N., Schaller, R.R.: Science and technology roadmaps. IEEE Transactions on engineering management. 48(2), 132-143 (2001)
14. Phaal, R., Farrukh, C., Probert, D.: Customizing roadmapping. Research-Technology Management. 47(2), 26-37 (2004)
15. Denner, M.S., Püschel, L.C., Röglinger, M. How to exploit the digitalization potential of business processes. Business & Information Systems Engineering. 60(4), 331-349 (2018)
16. Braun, C., Wortmann, F., Hafner, M., Winter, R.: Method Construction - a Core Approach to Organizational Engineering. In: Proceedings of the 2005 ACM Symposium on Applied Computing, pp. 1295–99. ACM, New York, NY, USA (2005)
17. Myers, M.D., Newman, M.: The qualitative interview in IS research: examining the craft. Inf Organ. 17(1), 2–26 (2007)
18. Gartner magic quadrant, https://www.gartner.com/en/research/methodologies/magic-quadrants-research (Accessed: 13.09.2018)
19. Alt, R., Beck, R., Smits, M.T. FinTech and the transformation of the financial industry. Electronic Markets. 28(3), 235-243 (2018)
20. Iansiti, M., Lakhani, K.R.: Managing Our Hub Economy: Strategy, Ethics, and Network Competition in the Age of Digital Superpowers. Harvard Business Review. 95(5), 84–92 (2017)
21. Mackenzie, A.: The Fintech Revolution. London Business School Review. 26(3), 50–53 (2015)

The Changing Roles of Innovation Actors and Organizational Antecedents in the Digital Age

Katharina Drechsler[1], Victoria Reibenspiess[1], Andreas Eckhardt[1], Heinz-Theo Wagner[1]

[1] German Graduate School of Management and Law, Heilbronn, Germany
{katharina.drechsler, victoria.reibenspiess, andreas.eckhardt, heinz-theo.wagner}@ggs.de

Abstract. Despite being acknowledged for playing a pivotal role in facilitating innovations in the digital age, there is a lack of research on the multifaceted role of digital innovation actors. This paper provides a systematic, multi-disciplinary literature review on innovation actors in a digital and non-digital context. Based on a search of 149 high-quality journals and conference proceedings, we identified 110 articles as relevant and categorized as well as synthesized the knowledge on innovation actors' role and organizational antecedents in a digital and non-digital context. We find an increasing focus on innovation actors' role in user communities in a digital context. Moreover, literature on organizational antecedents puts a stronger emphasis on allocating resources to innovation actors outside the organization. By analyzing extant research we provide a comprehensive summary on current knowledge and outline opportunities for future research on digital innovation actors.

Keywords: Digital Innovation Actors, Organizational Antecedents, Literature Review, Digital Innovation, Digital Age

1 Introduction

Digital technology has given rise to a radically new type of innovation [1]. These digital innovations have been conceptualized as "carrying out new combinations of digital and physical components to produce novel products" [2, p.725]. The transformation in the nature of innovations' outcome has not only manifested itself in more heterogeneous knowledge [3], and an increased importance of network effects [1] but has also given rise to a more distributed agency [4]. Consequently, a new set of digital innovation actors with distinct proficiencies has emerged [1].

The innovation management literature has acknowledged the importance of innovation actors early on by pointing out their key role in innovation development: "A new idea either finds a champion or dies" [5, p.84] and "successful innovation […] require a special combination of entrepreneurial, managerial and technical roles" [6, p.59]. Innovation actors are defined as stakeholders who promote an innovation vigorously through the various stages of the development process against resistance and by taking risks [5–8]. Existing reviews in innovation management [e.g., 7, 9]

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

have considered the concept of innovation actors through the perspective of their particular subdiscipline, without considering the new materiality of digital innovation.

In a digital context, a number of studies with various research foci have explored digital innovation actors' roles and organizational antecedents [e.g., 10, 11]. However, literature reviews that synthesize the current state of knowledge on digital innovation actors are very scarce so far [12]. Thus, existing reviews focus on digital innovation, but neglect to consider literature on innovation actors [e.g., 13, 14]. Moreover, no comprehensive literature reviews exist that explore fundamental differences in innovation actors' roles and organizational antecedents in a digital and non-digital context. Changes in innovation actors' roles and organizational antecedents that are caused by the distinct materiality of digital technology are unclear so far [1, 3]. With the rising importance of digital technology and the increasing prevalence of digital innovation such research is important. Organizations can only identify innovation actors and promote them by creating fitting organizational conditions, if innovation actors' roles in a digital and non-digital context are sufficiently clear [15].

Literature acknowledges this gap and called for future research to explore innovation actors' roles in a digital context by acknowledging "the complexity of how their actions interact with, and can be shaped by, a wider change process" [12, p.108]. Therefore, we explore the following two research questions:

RQ1: What are distinct roles of innovation actors in a digital and non-digital context?
RQ2: Which organizational characteristics promote or hinder innovation actors in a digital and non-digital context?

In a nutshell, this research article provides a comprehensive literature review on innovation actors' roles and organizational antecedents in a digital and non-digital context. By presenting an in-depth analysis of four subdisciplines and synthesizing findings from an individual and organizational perspective, this literature review offers the opportunity to build a thorough understanding of innovation actors. Based on differences in digital and non-digital innovation literature, we also identify gaps in existing research and provide practical implications.

The paper is structured as follows. While the next section outlines the methodology, section 3 describes the findings of our content-based analysis. Next, we discuss our results with implications for theory and practice and identify avenues for future research. Finally, we delineate our study's limitations.

2 Methodology

With respect to the methodology, a narrative literature review [16] was performed following a systematic and transparent methodology based on Paré et al. [17]. For the purpose of assuring the findings' quality, our search process compromised six steps adopted from Rowe [18]: selecting research questions, choosing sources, creating a search string, applying methodological and practical screening criteria, categorizing and reviewing literature's findings as well as synthesizing the results.

First, we selected a research question (see section 1). In a second step, we chose the sources for our literature search by opting for leading journals in four subdisciplines, information systems, organization and human resources, business administration as well as technology, innovation and entrepreneurship to account for the interdisciplinary nature of the research theme. A meta-ranking (Journal Quality List [19]), which incorporates 12 different journal rankings (e.g., Financial Times 50 Ranking 2016 or German VHB-JOURQUAL3), was used to evaluate the publication outlets. The 149 selected publication outlets were classified as leading journals in the majority of these rankings and include among other outlets the AIS Senior Scholars' Basket of 8. When considering, for instance, the German VHB-JOURQUAL3 we included all journals, ranked in the categories A+, A or B. The literature search was restricted to the time frame 1995 to 2018, because the year 1995 marks the beginning of the Internet commercialization, characterized by the elimination of the last restrictions on its commercial use [20]. This acknowledges innovation actors' high importance for digital innovation development [21] and enables us to draw a comparison between a digital and non-digital context.

In a third step, we created and utilized an extensive search string within the selected journals including five keywords: innovation, championing, level of analysis, characteristics and context. As depicted in Table 1 each keyword was covered by a variety of search terms, including synonyms as well as corresponding adjectives and verbs. To cover innovation actors comprehensively as well as systematically and to limit prepossessions on the research topic, we chose a broad range of synonymous and overlapping search terms to characterize innovation actors who promote innovation. We consolidated activities, such as brokering and promoting to cover innovation actors' roles in both a non-digital (e.g., innovation champions [8]) and digital context (e.g., lead users [10]). Moreover, we consider innovation actors both from an individual and organizational perspective.

Table 1. Search string

Keyword	*Search terms*
Innovation	("innovat*")
Championing	("champion*" OR "promot*" OR "boundary spann*" OR "broke*" OR "recombin*" OR "cataly*" OR "sponsor*" OR "corporate entrepreneur*" OR "blog*" OR "challeng*" OR "use*" OR "develop*" OR "influenc*")
Level of Analysis	("individ*" OR "personal*" OR "user*" OR "human" OR "employee") OR ("organi?ation*" OR "network*")
Characteristics	("characteristic*" OR "behav*" OR "attribute*" OR "trait*" OR "propert*" OR "qualit*" OR "capabilit*" OR "structure*" OR "culture*" OR "factor*" OR "requirement*" OR "variable*" OR "element*" OR "competence*" OR "nature*" OR "personalit*")
Context	"digital"

At least one search term related to each keyword had to appear either in the title, the abstract or the subject terms in order to be considered relevant for our literature review. Accordingly, a complete search string was generated. Next, a literature search was executed by using this search string and a meta-search engine, based on 202 different databases, such as EBSCO Business Source Complete, and containing all relevant 149 publication outlets. We ran our search both with and without the search term “digital” to cover both a digital and non-digital context. In the search, 1178 research articles were identified as potentially relevant.

In a fourth step, we screened the potentially relevant research articles grounded on five methodological and practical criteria. The examined articles had to (1) include a research methodology, (2) address aspects of the innovation process, (3) analyze an actor championing innovation, (4) adopt an individual or organizational perspective, and (5) not focus on the macro level. First these filtering criteria were applied to the title, abstract and keywords resulting in a reduction of the relevant research articles to 270. Second, the full text was screened, which led us to 85 relevant research articles. Following Webster and Watson [22], we then performed a backward (i.e., reviewing older literature quoted in the relevant papers) and a forward search (i.e., reviewing sources that quoted the article) to include all literature sources on innovation actors, which resulted in 25 additional research articles. Overall, the final sample consisted of 110 relevant papers.

Fifth, we categorized and reviewed literature’s findings. We covered innovation actors both from an individual and organizational perspectives by categorizing the content of the research articles into (1) (digital) innovation actors’ roles and (2) six dimensions of organizational antecedents. The different roles of innovation actors were derived in an iterative and inductive process. To analyze literature on organizational antecedents systematically, we adopted a categorization from prior research [23–25] that has found these six organizational characteristics[1], structure [24], human resource (HR) practices [25], culture & climate [24], resource allocation [24], knowledge management [24], and strategy [24], to influence organizational innovativeness. Finally, we synthesized literature’s findings, as elaborated in the following section (step 6).

3 Analysis

In the following, we first analyze innovation actors’ roles (individual perspective). Next, we outline our findings on organizational antecedents, characteristics that facilitate or hinder innovation actors’ innovativeness (organizational perspective). Figure 1 illustrates our theoretical framework based on our categorization schema.

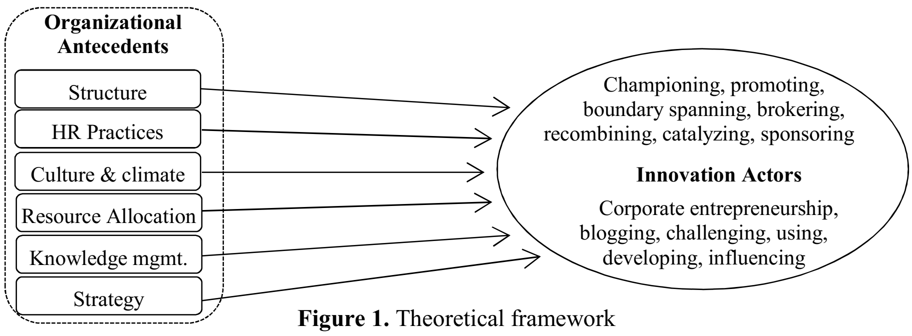


Figure 1. Theoretical framework

[1] A table depicting the definitions of the different categories can be accessed here: https://bit.ly/2ToX7cS

3.1 Roles of Innovation Actors

Innovation Actors in a non-digital context. Our analysis showed that we can distinguish between a number of innovation actors in a digital and non-digital context. The innovation champion constitutes one of the earliest identified innovation actors [5], who promotes an innovation vigorously through the various stages of the development process against potential resistance by taking risks [e.g., 6, 8]. Innovation champions have been described as motivating their innovation team [e.g., 26], inspiring others with their vision [e.g., 15], transferring information and knowledge [e.g., 27], connecting with others and building networks [e.g., 26], bringing different actors in the organization together [e.g., 28] and gaining management support [e.g., 8]. Besides the innovation champion, further types of innovation actors[2] have been identified as summarized in the following table.

Table 2. Roles of Innovation Actors in non-digital and digital context

Roles of Innovation Actors	*Exemplary Sources*
Innovation champion – Innovation actor who promotes an innovation vigorously through the various stages of the development process against potential resistance by taking risks. [6, 27] (Synonym: *process promoter*)	*Non-digital:* [8, 28] *Digital:* [29]
Corporate entrepreneur – Innovation actor who creates a new venture or initiates renewal or innovation within an existing organization by combining four competencies: inventing, brokering, championing and sponsoring. [30]	*Non-digital:* [31, 32] *Digital:* [33, 34]
Sponsor – Innovation actor who holds a managerial position and uses his or her formal power to support an innovation by supplying or obtaining resources, lending legitimacy or giving advice. [6, 27] (Synonym: *power promoter*)	*Non-digital:* [30, 35] *Digital:* [11, 36]
Boundary spanner – Innovation actor who is responsible for the interaction of an organizational unit or organization with its environment. [27, 37] (Synonym: *relationship promoter*)	*Non-digital:* [35, 37] *Digital* [38]
Knowledge broker – Innovation actor who facilitates information flows by transferring knowledge important in the innovation process between otherwise unconnected actors. [39]	*Non-digital:* [39, 40] *Digital:* [41]
Lead user – Innovation actors on the user side who detects problems, generates ideas for improvements to existing products and subsequently carries out modifications to generate an innovative product. [10, 42]	*Non-digital:* - *Digital:* [43, 44]

Innovation Actors in a digital context. The roles of innovation actors described above are also mentioned in a digital context. At the same time, digital technology gives rise and puts special emphasis on two roles, lead users and sponsors. Lead users have been shown to drive innovations from a user perspective in a digital context [e.g., 10, 43]. They communicate and collaborate with other (lead) users in user communities via digital platforms or technologies and apply their own knowledge and

[2] Even though roles of innovation actors have been characterized extensively in the literature, we only focus on the most frequently mentioned activities characterizing innovation actors. Therefore, the cited references only represent a selection of the research articles that we considered in the analysis overall.

knowledge exchanged with other users to advance products and drive innovation [e.g., 42]. A purely user-specific role of innovation actors has only been enabled by the distinct characteristics of digital technology. Moreover, literature in a digital context puts a stronger emphasis on the role of sponsors. Especially the new organizational role of the chief digital officer, one type of sponsor, has gained considerable significance, as these innovation actors in management positions drive an organization's digital transformation and champion digital innovation [e.g., 11].

3.2 Organizational Antecedents

We now analyze how idiosyncratic characteristics of an organization influence and shape innovativeness of digital and non-digital innovation actors following the framework depicted in Figure 1. All findings are summarized in Table 3[3].

Structure. As an organizational antecedent to innovation actors' effectiveness in a non-digital context organizational structure has been widely studied. A high degree of centralization in decision-making and a high degree of formalization of behavior through rules and procedures have been found to form barriers for non-digital innovation actors' emergence and effectiveness [e.g., 37, 45]. In contrast, a low degree of vertical differentiation, i.e. the existence of few hierarchical levels, and structuring an organization into teams and based on projects, an aspect of horizontal differentiation, enhances non-digital innovation actors' activities [e.g., 30].

These findings for the non-digital context are in line with evidence found in a digital context. Thus, Ansari and Munir [43] find that organizations need to move from a structure characterized by hierarchy and control to a structure that enables collaborative and interactive innovation with digital innovation actors in user communities. In addition, digital innovation research also focuses on other aspects of organizational structure. For instance, digital innovation actors in high hierarchical positions can only champion innovation effectively if their role is defined clearly and in alignment with other executive positions [e.g., 11, 36].

HR Practices. With respect to HR practices, research in a digital or non-digital context shows similar findings but addresses distinct types of innovation actors, respectively. When focusing on performance appraisal, sanctions due to failed innovation projects are likely to impede the emergence of non-digital innovation actors. Contrarily, rewards compensating innovation actors for innovation success enhance their emergence in a non-digital context [e.g., 32, 45]. However, non-digital literature disagrees whether performance appraisal should be based on innovation-promoting behavior [30] or innovation accomplishments [45]. Digital innovation research finds that performance appraisal needs to also compensate digital innovation actors in user communities for innovation success monetarily [e.g., 46] or through non-financial measures, such as recognition or rewards [e.g., 10, 44] to enhance

[3] In Table 3 different organizational characteristics are considered from the perspective of drivers only. As elaborated, for some of these factors only the counterfactual relations with innovation actors have been explored. In the illustration in Table 3 we inversely code barriers in order to display drivers.

digital actors' activities. Additionally, Tumbas et al. [11] point out the importance of defining key performance indicators for all executives driving digital innovation.

On staffing practices non-digital innovation literature remains largely silent and only proposes hiring employees with distinct personalities [e.g., 30] to spur innovation actor's emergence. In a digital context, hiring employees with high experience inside and outside the organization is found to be positively associated with innovation-promoting behavior [e.g., 41]. When considering a group of digital innovation actors that work together to advance an organization's innovation projects, Van Laere and Aggestam [29] propose that a diverse group of individuals who possess complementary skills, knowledge, and social networks should be hired to enhance digital innovation actors' effectiveness. Additionally, in digital innovation processes that incorporate a user community hiring leaders of the community as gatekeepers between community and organization can enhance innovation promotion within the user community, because these gatekeepers moderate the exchange and simultaneously maintain the boundaries between community and firm [e.g., 44].

Training employees is another aspect of HR practices that has been shown to be positively associated with innovation actors' emergence and effectiveness in both research streams [e.g., 31, 37]. While non-digital literature generally focuses on employees, in a digital context the training of external users is also beneficial for the emergence of digital innovation actors [e.g., 44].

Culture & Climate. Non-digital literature finds that a culture supportive towards innovation [31, 32] is positively associated with innovation actors' activities. Both digital and non-digital innovation research agree that a long-term outcome orientation of the business culture [e.g., 30, 34] as well as culture tolerant of failure [e.g., 10, 31] and risk rewarding [e.g., 32, 47] encourage (digital) innovation actors' emergence and effectiveness. Additionally, in a digital context organization's culture needs to adapt to external users' participative role in the innovation process to encourage the emergence of innovation actors in user communities [e.g., 43]. Going one step further, Parmentier and Mangematin [44] find that organizations need to work towards identity convergence of user community and organization by sharing identifying elements and building common values embedded in products and services.

Resource Allocation. The non-digital literature on resource allocation's influence on innovation actors presents a positive effect of provisioning financial resources and time to pursue innovation [e.g., 32], as well as management legitimization to use existing resources or networks [e.g., 47]. In a digital context, the availability of digital technology can promote, and limitations to technological capabilities can hinder innovation actors' promotion of innovation [e.g., 33, 48]. If no formal allocation of resources towards digital innovation occurs, a lack of internal control benefits digital actors' effectiveness, because it allows the diversion of funds and employees [e.g., 34]. Additionally, literature on digital innovation emphasizes allocating resources to the innovating user community. The provision of tools for innovation-promoting activities as well as support towards the community (e.g., through community events) can enhance digital innovation actors' emergence [e.g., 10, 44].

Table 3. Organizational antecedents of non-digital and digital innovation actors

Category	*Non-digital innovation actors*	*Digital innovation actors*
Structure	Low centralization and formalization [e.g., 37, 45] Low vertical differentiation [e.g., 30] Horizontal differentiation into teams [e.g., 30]	Structure enabling collaborative and interactive innovation [e.g., 43] Definition and alignment of role on executive level [e.g., 11, 36]
Human Resource Practices	Existence of rewards, but no sanctions [32, 45] Performance appraisal based on innovation-promoting behavior [e.g., 30] or innovation accomplishment [e.g., 45]	Monetary compensation [e.g., 46] or non-financial rewards for digital innovation actors in user community [e.g., 10, 44] Definition of key performance indicators on executive level [e.g., 11]
	Hiring employees with distinct personalities [e.g., 30]	Hiring employees with high experience inside and outside organization [e.g., 41] Hiring diverse group of individuals with complementary skills, knowledge and social networks [e.g., 29] Hiring leaders of community as gatekeepers [e.g., 44]
	Training of employees [e.g., 31, 37]	Training of external users [e.g., 44]
Culture & climate	Culture supportive towards innovation [e.g., 31, 32] Long-term outcome orientation [e.g., 30] Culture tolerant of failure and risk rewarding [e.g., 31, 32, 47]	Long-term outcome orientation [e.g., 34] Culture tolerant of failure and risk rewarding culture [e.g., 10] Adaption of culture to users' participative role [e.g., 43] Identity convergence of user community and organization [e.g., 44]
Resource allocation	Provision of financial resources and time [e.g., 32] Management legitimization to use existing resources and networks [e.g., 47]	Availability of digital technology and technological capabilities [e.g., 33, 48] Lack of internal control allowing the diversion of funds [e.g., 34] Allocation of resources to user community [e.g., 10, 44]
Knowledge Mgmt.	General learning orientation of organization [e.g., 31] Organizational support towards knowledge exploitation and recombination [e.g., 49]	Tools and databases supporting the sharing, exchange and creation of knowledge [e.g., 10, 50] Creation of interaction possibilities in user community [e.g., 10, 44] Sharing knowledge with external users [e.g., 10, 44]
Strategy		Effective IT governance structure [e.g., 36] Opening content to user community without losing control [e.g., 43, 44, 46]

Knowledge Management. With respect to knowledge management, non-digital literature is relatively silent and only proposes that a general learning-orientation in organizations [e.g., 31] and organizational support towards knowledge exploitation and recombination [e.g., 49] strengthen innovation actors' emergence and effectiveness. In a digital context, tools and databases that support the exchange and creation of knowledge promote digital innovation actors' effectiveness [e.g., 10, 50]. In innovation processes incorporating a user community, organizations can promote digital innovation actors' effectiveness by creating multiple possibilities of interaction (e.g., setting up discussion areas) to foster the exchange of explicit and tacit knowledge and by actively sharing knowledge [e.g., 10, 44].

Strategy. While non-digital literature is relatively silent on the role of strategy for innovation actors' emergence and effectiveness, digital technology poses new challenges that need to be addressed. Thus, for digital innovation actors on the executive level an effective information technology (IT) governance structure is a requirement for their effectiveness [e.g., 36]. In innovation processes involving a user community, organizations' optimal strategy to promote digital innovation actors' emergence incorporates opening (proprietary) content [e.g., 46] without losing control of the innovation outcome [e.g., 43, 44].

4 Discussion and Areas of Future Research

This research offers a comprehensive literature review on differences in innovation actors' roles (RQ1) and organizational antecedents (RQ2) in a digital and non-digital context. Theoretically, we contribute to literature by providing an in-depth analysis of research in four subdisciplines. By contrasting findings on innovation actors in a digital and non-digital context and taking both an individual and organizational perspective, our literature review offers insights into changes caused by the distinct materiality of digital technology and aims to close the identified gap in literature [12]. Based on our findings, we also provide recommendations and research questions for promising avenues of future research (see below).

Practically, we contribute to literature by offering organizations' management important insights into changes in innovation actors' roles due to the digital transformation. By synthesizing differences in organizational antecedents in a digital and non-digital context, we also enable organizations to provide adequate framework conditions to support innovation actors and enable the championing of innovation.

4.1 Roles of Innovation Actors

With regard to the roles of innovation actors (see subsection 3.1), our analysis shows that most roles of innovation actors hardly vary in a digital compared to a non-digital context. At the same time, research points to the rise of a new role, lead users in innovation collectives (i.e., user communities), and puts a higher emphasis on one role already known in a non-digital context: sponsors (e.g., chief digital officers) [11, 43]. The reason and importance of these changes remain unclear throughout existing literature. Furthermore, we observe that research on digital innovation actors' roles is rare, indistinct and ambiguous. Literature in a digital context rarely provides a characterization of innovation actors that goes beyond a description of innovation actors' behaviors and incorporates their knowledge, skills and personality profile.

Additionally, innovation actors show different degrees of homogeneity in a digital and non-digital context. Thus, innovation actors' roles in a non-digital context are characterized by similar behaviors and share common objectives (i.e., innovation champion and corporate entrepreneur) [27]. In contrast, innovation actors' roles in a digital context vary more greatly. While sponsors in a digital context, such as chief digital officers, are part of the management board [11], lead users can rarely influence organizations' strategic decisions [42]. At the same time, these heterogeneous, digital innovation actors are increasingly part of a group or innovation community [29, 42]. Since, the scarcity of research on groups of innovation actors in a non-digital context limits the implications that can be derived for a digital context, the characteristics and compositions of such groups offer another area for future research. By considering and combining the results on innovation actors' role, Table 4 integrates and concludes with recommendations for future research and proposes research questions.

Table 4. Research agenda for future research on innovation actors' roles[4]

Recommendation	*Selected research questions for future research*
Researchers should investigate innovation actors' roles in a digital context	**Why do new roles, such as lead users in innovation collectives, arise and why does the emphasis with respect to existing roles change?** How important are these changes for digital innovation?
	What characterizes digital innovation actors' knowledge, skills and personality?
	How do digital innovation actors' goals and motivation differ from those of non-digital innovation actors due to the rise of digital technology?
	How do innovation actors develop the skillset required for digital innovation?
Researcher should analyze characteristics and compositions of groups of digital innovation actors	**How can groups of innovation actors as well as their composition be characterized?**
	What are the factors enabling or hindering the collaboration of digital innovation actors in a group?
	Does artificial intelligence change the collaboration of digital innovation actors in innovation communities?

4.2 Organizational Antecedents

With regard to the organizational characteristics that enable or hinder innovation actors, our results reveal changes in organizational antecedents associated with the distinct materiality of digital technology. We find that literature focuses on different aspects of innovation actors' organizational antecedents in a digital compared to a

[4] All research questions in Table 4 and 5 were derived from the literature review's findings. Research questions in bold are discussed in detail in the respective subsection.

non-digital context. Digital innovation literature puts a stronger focus on the inclusion of innovation actors outside the organization (i.e., lead users), for example, by allocating resources and training these actors. Moreover, only a small number of articles examine organizational antecedents in a digital context [e.g., 33, 46]. These papers remain very generic and have mainly other research foci, so that organizational antecedents are only covered shallowly. Since organizational antecedents have been shown to play such an important role in enabling organizational innovativeness [e.g., 24], future research should explore a number of aspects in depth.

The existing literature on organizational antecedents in a digital context points in one direction: the digitization of work environments initiates change, which rapidly redefines the interaction of individuals and organizations. We observe that organizational boundaries are weakened as user communities play an increasingly important role. For instance, with respect to the organizational structure, literature finds that a structure enabling collaborative and interactive innovation not only inside the organization but also in user communities outside the organization is beneficial for digital innovation actors [43]. Similarly, in a digital context organizations not only need to create a culture internally but also have to establish a shared culture with the user community [44]. In the future, digital technology could lead to the dissolution of traditional organization structures towards virtual organizations with a loose accumulation of innovation actors and new forms of collaboration between them [1]. Challenges connected to these developments have not been addressed in existing literature so far.

Issues could, for instance, arise with respect to resource allocations to digital user communities. A virtual organization would not only need to provide appropriate IT infrastructure and resources to the digital user community, but also ensure the correct and targeted usage. Yet, the tracking of resources to secure efficient usage in a digital environment could prove to be more challenging as innovation actors would be scattered all around the world. Moreover, with the increasing heterogeneity of the innovation actors, the individual requirements to IT infrastructure might diverge [4], further adding to the challenge. Therefore, questions on how to allocate and use resources efficiently to avoid the waste of resources need to be explored in-depth in the future. Similarly, if innovation actors collaborate with organizations spontaneously using digital platforms in user communities, it will be difficult to track their knowledge and skills. Due to the nature of platforms, innovation actors will vary and their composition fluctuate [51]. As a result, knowledge becomes more tacit and fluid [3]. To face this challenge new knowledge management systems have to be created and new avenues for future research exist.

While we have discussed on a limited number of potential avenues for future research in the following, Table 5 integrates and concludes with more elaborate recommendations and research questions for future research on organizational antecedents.

Table 5. Research agenda for future research on innovation actors' organizational antecedents

Recommendation	*Selected research questions for future research*
Researchers should investigate innovation actors' organizational antecedents in a digital context:	**Why do organizational antecedents promoting or hindering digital innovation actors change due the rise of digital technology?** Do organizational antecedents that promote or hinder non-digital innovation actors also affect digital innovation actors? Which additional organizational factors could hinder the evolution and development of digital innovation actors?
Structure	**Which structure is required to enable collaborative and interactive innovation among digital innovation actors not only inside the organization but also in user communities outside the organization?** How can organizations establish such a structure?
Human Resource Practices	What are appropriate incentives to motivate digital innovation actors? How can HR departments identify digital innovation actors? How can non-digital innovation actors evolve into digital innovation actors? How can organizations support non-digital innovation actors in this endeavor?
Culture & climate	**How can organizations create a culture that supports digital innovation actors both inside and outside an organization in their endeavor to promote innovation?**
Resource allocation	**What are the requirements for appropriate IT infrastructure and resource allocation to digital innovation actors in virtual organizations?** **How can organizations promote efficient use of resources among heterogeneous innovation actors in innovation communities?** What are the diverging requirements of digital innovation actors with respect to IT infrastructure and resources?
Knowledge Management	**What are the requirements for organizations' knowledge management systems to track knowledge and skills of digital innovation actors in user communities?** How can organizations ensure an appropriate and complementary composition of innovation actors' knowledge and skills in user communities?
Strategy	How can organizations manage the strategic challenge of opening proprietary content to digital innovation actors in innovation communities without losing control of innovation outcomes?

5 Limitations of the Literature Review

After the preceding analysis and discussion of our findings we also acknowledge some limitations. The selection of publications of our review restricts the results of our analysis. The review is based on 149 publication outlets selected using a meta-ranking (Journal Quality List [19]), covering 12 different journal rankings. Although this selection ensures the high quality of our literature base, some relevant contributions, such as scientific books [e.g., 52] or whitepapers, may be missing in the review due to the restriction of our sample to peer-reviewed publications. Similarly, by limiting the time frame of our search to 1995 to 2018 we risk the exclusion of relevant literature. Since the concept of the innovation actors was first mentioned in 1963 [5], relevant research articles might have been published prior to 1995. However, we solve this problem by relying on backward search to complement our sample of the relevant literature [22].

Moreover, the coding and categorization of innovation actors' roles and organizational antecedents may have been subject to mistakes. Yet, since we relied on two independent coders, who followed an orderly and rigorous coding approach, the number of mistakes was kept to a minimum. Accordingly, a high reliability and validity of the findings of our analysis was secured [53].

References

1. Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M.: Digital innovation management: Reinventing innovation management in a digital world. MIS Q. 41, 223–238 (2017).
2. Yoo, Y., Henfridsson, O., Lyytinen, K.: The new organizing logic of digital innovation: An agenda for information systems research. Inform. Syst. Res. 21, 724–735 (2010).
3. Yoo, Y., Boland, R.J., Lyytinen, K., Majchrzak, A.: Organizing for innovation in the digitized world. Organ. Sci. 23, 1398–1408 (2012).
4. Fichman, R.G., Dos Santos, B.L., Zheng, Z.: Digital innovation as a fundamental and powerful concept in the information systems curriculum. MIS Q. 38, 329–353 (2014).
5. Schon, D.: Champions for radical new inventions. Harv. Bus. Rev. 41, 77–86 (1963).
6. Maidique, M.A.: Entrepreneurs, champions and technological innovation. Sloan Manag. Rev. 21, 59–76 (1980).
7. Jenssen, J.I., Jørgensen, G.: How do corporate champions promote innovations? Int. J. Innovat. Manag. 8, 63–86 (2004).
8. Howell, J.M., Shea, C.M.: Individual differences, environmental scanning, innovation framing, and champion behavior: key predictors of project performance. J. Prod. Innovat. Manag. 18, 15–27 (2001).
9. Elkins, T., Keller, R.T.: Leadership in research and development organizations: A literature review and conceptual framework. Leader. Q. 14, 587–606 (2003).
10. Mahr, D., Lievens, A.: Virtual lead user communities: Drivers of knowledge creation for innovation. Res. Pol. 41, 167–177 (2012).
11. Tumbas, S., Berente, N., Brocke, J. vom: Digital innovation and institutional entrepreneurship: Chief Digital Officer perspectives of their emerging role. J. Inform. Tech. 33, 188–202 (2018).
12. Holmström, J.: Recombination in digital innovation: challenges, opportunities, and the importance of a theoretical framework. Inform. Organ. 28, 107–110 (2018).
13. Kohli, R., Melville, N.P.: Digital innovation: A review and synthesis. Inform. Syst. J. 1–24 (2018).
14. Barrett, M., Davidson, E., Prabhu, J.: Service innovation in the digital age: key contributions and future directions. MIS Q. 39, 135–154 (2015).
15. Howell, J.M., Higgins, C.A.: Champions of technological innovation. Admin. Sci. Q. 35, 317–341 (1990).

16. Paré, G., Tate, M., Johnstone, D., Kitsiou, S.: Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. Eur. J. Inform. Syst. 25, 493–508 (2016).
17. Paré, G., Trudel, M.C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: A typology of literature reviews. Inform. Manag. 52, 183–199 (2015).
18. Rowe, F.: What literature review is not: diversity, boundaries and recommendations. Eur. J. Inform. Syst. 23, 241–255 (2014).
19. Harzing, A.-W.: Journal quality list, http://www.harzing.com/resources/journal-quality-list (Accessed: 05.03.2018).
20. Harris, S.R., Gerich, E.: Retiring the NSFNET backbone service: chronicling the end of an era. ConneXions. 10 (1996).
21. Beath, C.M.: Supporting the information technology champion. MIS Q. 15, 355–372 (1991).
22. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a review. MIS Q. 26, xiii--xxiii (2002).
23. Smith, M., Busi, M., Ball, P., Van der Meer, R.: Factors influencing an organisation's ability to manage innovation: a structured literature review and conceptual model. Int. J. Innovat. Manag. 12, 655–676 (2008).
24. Crossan, M.M., Apaydin, M.: A multi-dimensional framework of organizational innovation: a systematic review of the literature. J. Manag. Stud. 47, 1154–1191 (2010).
25. Chen, C.-J., Huang, J.-W.: Strategic human resource practices and innovation performance - the mediating role of knowledge management capacity. J. Bus. Res. 62, 104–114 (2009).
26. Howell, J.M., Shea, C.M.: Effects of champion behavior, team potency, and external communication activities on predicting team performance. Group. Organ. Manag.. 31, 180–211 (2006).
27. Chakrabarti, A.K., Hauschildt, J.: The division of labour in innovation management. R&D Manag. 19, 161–171 (1989).
28. Shane, S., Venkataraman, S., MacMillan, I.: Cultural differences in innovation championing strategies. J. Manag. 21, 931–952 (1995).
29. van Laere, J., Aggestam, L.: Understanding champion behaviour in a health-care information system development project – how multiple champions and champion behaviours build a coherent whole. Eur. J. Inform. Syst. 25, 47–63 (2016).
30. Hayton, J.C., Kelley, D.J.: A competency-based framework for promoting corporate entrepreneurship. Hum. Resource Manag. 45, 407–427 (2006).
31. Anthony, S.D.: The new corporate garage. Harv. Bus. Rev. 90, 44–53 (2012).
32. Hornsby, J.S., Kuratko, D.F., Zahra, S.A.: Middle managers' perception of the internal environment for corporate entrepreneurship: assessing a measurement scale. J. Bus. Vent. 17, 253–273 (2002).
33. Austin, R.D., Devin, L., Sullivan, E.E.: Accidental innovation: Supporting valuable unpredictability in the creative process. Organ. Sci. 23, 1505–1522 (2012).
34. Abetti, P.: The birth and growth of Toshiba's laptop and notebook computers: a case study in Japan. J. Bus. Vent. 12, 507–529 (1997).
35. Glaser, L., Fourné, S., Elfring, T.: Achieving strategic renewal: the multi-level influences of top and middle managers' boundary-spanning. Small Bus. Econ. 45, 305–327 (2015).
36. Peppard, J., Edwards, C., Lambert, R.: Clarifying the ambiquous role of the CIO. MIS Q. E. 10, 115–117 (2011).
37. Ebers, M., Maurer, I.: Connections count: How relational embeddedness and relational empowerment foster absorptive capacity. Res. Pol. 43, 318–332 (2014).

38. Harryson, S.: Entrepreneurship through relationships–navigating from creativity to commercialisation. R & D Manag. 290–310 (2008).
39. Kirkels, Y., Duysters, G.: Brokerage in SME networks. Res. Pol. 39, 375–385 (2010).
40. Belso-Martínez, J.A., Molina-Morales, F.X., Martínez-Cháfer, L.: Contributions of brokerage roles to firms' innovation in a confectionery cluster. Tech. Anal. Strat. Manag. 27, 1014–1030 (2015).
41. Ramirez, M., Dickenson, P.: Gatekeepers, knowledge brokers, and inter-firm knowledge transfer in Beijing's Zhongguancun Science Park. Int. J. Innovat. Manag. 14, 93–122 (2010).
42. Colazo, J.: Performance implications of stage-wise lead user participation in software development problem solving. Decis. Support Syst. 67, 100–108 (2014).
43. Ansari, S., Munir, K.: Letting users into our world: Some organizational implications of user-generated content. Res. Sociol. Org. 29, 79–105 (2010).
44. Parmentier, G., Mangematin, V.: Orchestrating innovation with user communities in the creative industries. Technol. Forecast. Soc. Change. 83, 40–53 (2014).
45. Globocnik, D., Salomo, S.: Do formal management practices impact the emergence of bootlegging behavior? J. Prod. Innovat. Manag. 32, 505–521 (2015).
46. Arakji, R., Lang, K.: Digital consumer networks and producer-consumer collaboration: Innovation and product development in the video game industry. J. Manag. Inform. Syst. 24, 195–219 (2007).
47. Halme, M., Lindeman, S., Linna, P.: Innovation for inclusive business: intrapreneurial bricolage in multinational corporations. J. Manag. Stud. 49, 743–784 (2012).
48. Panourgias, N.S., Nandhakumar, J., Scarbrough, H.: Entanglements of creative agency and digital technology: A sociomaterial study of computer game development. Technol. Forecast. Soc. Change. 83, 111–126 (2014).
49. Lichtenthaler, U., Ernst, H.: The role of champions in the external commercialization of knowledge. J. Prod. Innovat. Manag. 26, 371–387 (2009).
50. Flath, C.M., Friesike, S., Wirth, M., Thiesse, F.: Copy, transform, combine: Exploring the remix as a form of innovation. J. Inform. Tech. 32, 306–325 (2017).
51. Bogers, M., West, J.: Managing distributed innovation: Strategic utilization of open and user innovation. Creativ. Innovat. Manag. 21, 61–75 (2012).
52. Mansfeld, M.: Innovatoren: Individuen im Innovationsmanagement. Springer-Verlag, Wiesbaden (2011).
53. Tinsley, H.E., Weiss, D.J.: Interrater reliability and agreement of subjective judgments. J. Counsel. Psychol. 22, 358–376 (1975).

Bewertung des Kundennutzens von Chatbots für den Einsatz im Servicedesk

André Espig, Nicole Klimpel, Franz Rödenbeck und Gunnar Auth

Hochschule für Telekommunikation, Institut für Wirtschaftsinformatik, Leipzig, Deutschland
{andre.espig,nicole.klimpel,franz.roedenbeck,gunnar.auth}
@hft-leipzig.de

Abstract. Digitalisierung und fortwährend steigende Anforderungen an den Servicedesk führen dazu, dass zur Unterstützung bei der Kommunikation mit Kunden zunehmend autonome Software-Agenten, sog. Chatbots, zum Einsatz kommen. Mit dem dadurch steigenden Automatisierungsgrad lassen sich Effizienzverbesserungen und Einsparpotenziale realisieren. Allerdings hängt die Akzeptanz der Chatbots bei den Kunden von dem durch diese wahrgenommenen Nutzen ab. Im Rahmen dieser Arbeit wird eine Methode zur Bewertung des Kundennutzens von Chatbots im Bereich des Servicedesks entwickelt und evaluiert, die sich eines domänenbezogenen Fragekatalogs und spezifischer Bewertungskriterien bedient. Dabei folgt die Arbeit dem konstruktionsorientierten Forschungsansatz mittels Design Science Research Process und konzentriert sich auf die Bewertung von Chatbots aus Kundenperspektive.

Keywords: Servicedesk, Chatbot, Kundennutzen, Bewertungsmethode, Evaluierung

1 Einleitung

Im Zuge der Serviceorientierung im IT-Management hat sich der Servicedesk unternehmensintern als zentrale Kommunikationsschnittstelle zwischen IT-Organisation und ihren Nutzern weitgehend etabliert [1]. Als Vorbild dienten Organisationseinheiten im Kundenservice von Unternehmen. Diese haben die Aufgabe, den Kontakt mit externen Endkunden zu professionalisieren und dadurch die Kundenzufriedenheit zu verbessern [2]. Im Bestreben, Kundenanliegen schneller und effizienter zu bearbeiten, dabei zugleich aber auch Kosten zu reduzieren, setzen Unternehmen in letzter Zeit verstärkt sog. Chatbots ein [3]. Diese Software-Agenten zeichnen sich durch ihre (begrenzten) Fähigkeiten zur Verarbeitung natürlicher Sprache aus und ermöglichen dadurch eine weitere Automatisierung der Kunden- bzw. Nutzerschnittstelle [4]. Dabei stehen großen Erwartungen an die Automatisierungspotenziale durch Chatbots seitens der Unternehmen nicht weniger große Vorbehalte seitens der damit konfrontierten Zielgruppe gegenüber. So lehnen laut einer Studie aus dem Jahr 2017 [5] ca. 70 Prozent der Deutschen die neuen Sprachassistenten ab. Als eine mögliche Ursache für diese Ablehnung liegt eine mangelnde Nutzenstiftung durch Chatbots aus Kundensicht nahe. Neben dem Risiko der mangelnden Kundenakzeptanz wird Unterneh-

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

men das Erreichen ihrer Automatisierungsziele beim Einsatz von Chatbots auch durch das wachsende Angebot an Chatbot-Plattformen, Entwicklungsframeworks und Programmierschnittstellen sowie deren schnelle Weiterentwicklung zusätzlich erschwert [6]. Dadurch getrieben, sehen wir einen sich verstärkenden Bedarf, den tatsächlichen Kundennutzen von Chatbots festzustellen.

Ein Chatbot für den Einsatz im Servicedesk hat bestimmte Aufgaben zu erfüllen und Nutzerbedürfnisse zu befriedigen, die sich von denen anderer Chatbots (z.B. zur Wettervorhersage) unterscheiden. Die Evaluierung von Mensch-Computer-Dialogsystemen (hierzu zählen wir auch Chatbots) zielt seit dem Turing-Test von 1950 primär auf die Überprüfung der Intelligenz von Systemen mit Blick auf ihre Fähigkeit, Konversationen mit Menschen zu führen, ohne als Maschine erkannt zu werden [7], [8]. Aus Kundensicht ist diese Fähigkeit aber lediglich Mittel zum Zweck, nämlich durch Konversation mit einem Chatbot ein produkt- oder servicebezogenes Anliegen schnell und zufriedenstellend zu erfüllen.

Shawar und Atwell [9] haben verschiedene Evaluationsmethoden für Chatbots untersucht, die entweder aus Innensicht die Komponenten des Chatbots (Glass box evaluation) oder aus Außensicht den Chatbot als Ganzes (Black box evaluation) betrachten. Als Fazit wird betont, dass standardisierte Testverfahren ohne Berücksichtigung der jeweiligen Anwendungsdomäne und der damit verbundenen Kundenanforderungen wenig aussagekräftig seien. Insbesondere sollte sich die Bewertung nach der Lösungsqualität richten, die der Chatbot für eine ihm gestellte Aufgabe erzielt. Stoeckli et al. [6] untersuchen die Nutzenstiftung von Chatbots für abgegrenzte Nutzergruppen aus einer Functional-Affordance-Perspektive am Beispiel von Software-Entwicklungsteams. Functional Affordances (FA) repräsentieren die Möglichkeiten, die technische Artefakte bestimmten Nutzergruppen für deren Aufgabenverrichtung bieten. Damit abstrahiert dieser Bewertungsansatz gezielt von einzelnen Anwendungsfällen, um so grundsätzlichere Aussagen zum Chatbot-Nutzen treffen zu können. Dabei besteht ein enger Zusammenhang zwischen FA und den materiellen Eigenschaften des betrachteten Artefakts wie bspw. dessen konkrete Funktionalitäten (Features). Die eingenommene Perspektive stellt hier den Chatbot mit seinen feature-abhängigen Einsatzmöglichkeiten in den Mittelpunkt.

Demgegenüber nimmt der vorliegende Ansatz die Kunden- bzw. Nutzerperspektive ein und betrachtet den Chatbot als Blackbox, um so unabhängig von sich schnell verändernden technischen Implementierungen den vom Nutzer wahrgenommenen Mehrwert zu qualifizieren. Als weiteres Entwurfsziel für unsere Methode wird neben theoretischer Fundierung und hohem praktischen Nutzwert ein möglichst geringer Zeit- und Kostenaufwand für die Anwendung der Methode angestrebt. Einen vergleichbaren Evaluierungsansatz für Chatbots konnten wir in der bisherigen Literatur nicht finden.

Vor diesem Hintergrund geht unser Beitrag der Frage nach, wie sich der Kundennutzen eines Chatbots innerhalb der spezifischen Anwendungsdomäne Servicedesk systematisch bewerten lässt. Dabei liegt der Fokus auf Chatbots zur automatischen Beantwortung von Kundenanfragen. Als Ergebnis wird eine Bewertungsmethode vorgestellt, die durch eine fragebogengestützte Interaktion mit dem Chatbot Aussagen über dessen Nutzen aus der Perspektive des Kunden ermöglicht.

1.1 Abgrenzung der Begriffe Chatbot und Servicedesk

Der Begriff *Chatbot* (auch Conversational Agent [8] oder Virtual Assistant [10]) bezeichnet Software-Agenten, die in der Lage sind, mittels natürlicher Sprache Zug um Zug mit menschlichen Nutzern zu interagieren [9], also einen Dialog zu führen. Basierte die Intelligenz von Chatbots in der Vergangenheit meist auf regelbasierter Schlussfolgerung haben sich die Fähigkeiten durch neuartige Verfahren im Bereich maschinelles Lernen und Information Retrieval beträchtlich erweitert [11], [12]. Xu et al. [13] sehen in Chatbots eine vielversprechende Alternative zum herkömmlichen Kundenservice. Die Bedeutung von Chatbots im Kundenservice steigt durch ihr Potenzial als neuer, effizienter und automatisierter Kanal zur Kundeninteraktion, wobei Følstad et al. [14] auch darauf hinweisen, dass den Chatbots eine zunehmende Nutzung mobiler Messaging-Plattformen durch eine Vielzahl von Kundengruppen entgegenkommt.

Als Funktionseinheit einer Serviceorganisation besteht die Hauptaufgabe eines *Servicedesks* darin, den Kunden bei Problemen oder Anliegen im Zusammenhang mit bereitgestellten Services bzw. Produkte zu unterstützen [15]. Dazu fasst der Servicedesk sämtliche Mitarbeiter mit Support-Aufgaben organisatorisch zusammen und bietet den Kunden eine zentrale Kontaktstelle für Probleme und Anliegen aller Art (sog. Single point of contact). Für einen effizienten und effektiven Informationsaustausch mit den Kunden stehen typischerweise mehrere Kommunikationskanäle wie Telefon, E-Mail oder Web-Formulare zur Verfügung [1]. Damit werden eine effiziente und nachhaltige Bearbeitung der Kundenanliegen sowie eine möglichst schnelle Störungsbeseitigung angestrebt. Als „Gesicht der IT" zum Kunden liefert der Servicedesk einen wichtigen Beitrag zur Kundenzufriedenheit [16], [15]. Da der Servicedesk sehr personalintensiv ist, gerät er schnell unter Kostendruck. Gleichzeitig muss durch aktuelle Entwicklungen im Zuge von Digitalisierung und Industrie 4.0 für eine steigende Anzahl von Produkten und Services Support geleistet werden [1], [17]. Die Automatisierung des Servicedesks soll diese widersprüchlichen Anforderungen erfüllen. Schließlich erwarten die Benutzer ein immer besseres Nutzenerlebnis auf Basis eines wahrgenommenen Mehrwerts [18].

1.2 Forschungsmethodik

Der diesem Beitrag zugrundeliegende Forschungsprozess basiert auf einem konstruktionsorientierten Ansatz und folgt dem Design Science Research Process [19]. Nach Aufarbeitung des aktuellen Forschungsstands durch Literaturrecherche wurde als Artefakt eine Bewertungsmethode konstruiert. Die auf dem klassischen Turing-Test beruhende Loebner-Preis-Methode [20] diente dabei als Ausgangspunkt für eine weiterentwickelte, domänenbezogene Bewertungsmethode. Der jährlich vergebene Loebner-Preis, benannt nach dem Preis-Stifter Dr. Hugh Loebner, wurde 1991 erstmalig ausgelobt, um die Leistung von Software-Systemen in einem natürlichsprachlichen Mensch-Maschine-Dialog zu bewerten [21]. Beim Loebner-Preis handelt es sich um eine Blackbox-Evaluierung, für die seit 2014 ein standardisierter Fragebogen genutzt wird [20]. Die seit seiner ersten Durchführung immer wieder geäußerte Kritik am Loebner-Preis adressiert in erster Linie seine Eignung als Turing-Test [22], auch im Zusammenhang mit einer irreführenden Berichterstattung und Fehlinterpretationen in den Medien. Durch den Preis-Charakter besteht die Gefahr, dass durch einseitige Optimierungen und vordefinierte Reaktionen die allgemeine Aussagekraft der Bewertungsergebnisse verloren geht. Unabhängig von dieser Kritik eignet sich die Datenerhebung mittels standardisiertem Fragebogen für unsere Entwurfsziele und wird neben der grundlegenden Bewertungssystematik als Methodenelement übernommen.

Für die Konstruktion des Fragebogens wurden ausgehend von einer einfachen Konzeptionalisierung relevante Anforderungen für die Bewertung von Chatbots im Servicedesk aus der Literatur extrahiert und kategorisiert. Es folgte die Erarbeitung von Bewertungskriterien sowie eines Punktesystems zum Testen der Antwortqualität des Chatbot-Outputs. Parallel dazu wurden aus den Anforderungen Fragen für die Erstellung eines Fragenkataloges abgeleitet. Die Nutzung eines standardisierten Fragebogens ermöglicht die Erschließung der Kundenperspektive und eine gute Vergleichbarkeit der getesteten Chatbots. Aus den Bewertungskriterien und dem Fragenkatalog wurde ein Bewertungsbogen erstellt.

Die Entwicklung des Fragenkataloges beruhte zur Verbesserung der Objektivität auf der Anwendung einer Investigator-Triangulation, bei der nach getrennter Arbeitsphase die individuell entwickelten Fragestellungen inkrementell zu einem Gesamtkatalog vereint wurden. Mittels einer Testreihe mit fünf Chatbots erfolgte schließlich die Evaluation der Bewertungsmethode gemäß Design-Science-Prozess.

2 Entwurf einer Bewertungsmethode

Ähnlich wie Vasconcelos et al. [23] oder beim Loebner-Preis [20] wird die Annahme zugrunde gelegt, dass sich die Anforderungen an einen Chatbot dadurch testen lassen, dass ihm bestimmte Fragen gestellt werden. Der Chatbot wird unter der Einnahme der Kundenperspektive als Blackbox betrachtet, der ein bestimmter Input in Form von Fragen oder Bemerkungen übergeben wird und die daraufhin einen bestimmten Output generiert. Dieser vom Chatbot ausgegebene Output wird hinsichtlich seiner Qualität getestet und gibt so Aufschluss darüber, in welchem Ausmaß der Chatbot bestimmte Anforderungen erfüllt. Während der Loebner-Preis die Menschenähnlichkeit des Chatbots und Vasconcelos et al. [23] grundlegende Qualitäten der Gesprächsführung testen, strebt die hier entwickelte Methode die Bewertung eines domänenspezifischen Kundennutzens an. Damit unterscheidet sie sich in der eingenommenen Perspektive ebenfalls von der Bewertung des Chatbot-Nutzens für Programmierteams mittels Functional Affordances von Stoeckli et al. [6].

Die Konstruktion unserer Bewertungsmethode erfolgte in zwei Schritten: In einem ersten Schritt wurden ausgehend von einem grundlegenden Modell die für einen Chatbot im Servicedesk relevanten Anforderungen identifiziert. In einem zweiten Schritt erfolgte die Konzeption des eigentlichen Bewertungsvorgangs inklusive der Erstellung eines Bewertungswerkzeugs.

2.1 Identifikation wesentlicher Anforderungen

Die Grundlage für die Bewertungsmethode bildet das untenstehende Modell, welches das zu betrachtende System aus Kunde, Servicedesk und Chatbot beschreibt. Der Kunde äußert ein bestimmtes Anliegen, welches der Servicedesk aufnimmt und bearbeitet. Der Chatbot bildet das Interface zwischen Kunde und Servicedesk (siehe Abbildung 1).

Abbildung 1. Chatbot als Interface zwischen Kunde und Servicedesk

Das Vorgehen bei der Sammlung der Anforderungen an einen Chatbot für die Kundeninteraktion eines Servicedesks wird durch Abbildung 2 dargestellt. Dem Ausgangsmodell folgend, wurden Anforderungen aus den Zielen und Aufgaben des Servicedesks, aus allgemeinen Merkmalen kommerzieller Chatbots sowie aus den Kundenbedürfnissen abgeleitet. Die Anforderungen wurden in Anlehnung an Walker et al. [24] in die Kategorien Kommunikationsfähigkeit und Problemlösungsfähigkeit eingeordnet.

Um ein umfassendes Bild der Aufgaben im Servicedesk und ihrer Bedeutung für einen Chatbot zu erhalten, wurden die Anforderungen anhand der Komponenten des IT-Service-Managements – Personen, Prozesse, Technologie und Information – betrachtet [25]. Beispielsweise agiert der Chatbot an Stelle der Personen eines Servicedesks, so dass eine Bewertung daraufhin erfolgen sollte, wie gut er diese Aufgabe übernehmen kann. Wesentliche Aufgaben eines Servicesdesks (vgl. [16], [26], [27]) wurden den einzelnen Komponenten zugeordnet.

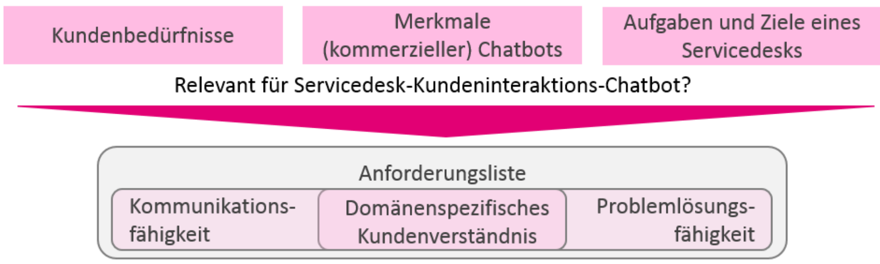


Abbildung 2. Ableitung der Anforderungen an einen Servicedesk-Chatbot

Anschließend wurden allgemeine Anforderungen an kommerzielle Chatbots [8] auf ihre Relevanz für Chatbots im Selfservice hin betrachtet. Hierzu gehört z. B. die Fähigkeit, mit Irritationen wie Rechtschreibfehlern oder Beleidigungen umzugehen sowie dem Kunden Feedbackmöglichkeiten anzubieten. Relevante Merkmale eines Chatbots aus dem Reifegradmodell von Smiers [28] wurden ebenfalls aufgenommen.

Um eine Bewertung des Kundennutzen ohne die Durchführung einer Nutzerbefragung zu ermöglichen, wurden die von Nielsen [29] entwickelten

Heuristiken in ihrer Übertragung auf Chatbots nach dem Vorschlag von Scott [30] verwendet. Demnach sollte ein Chatbot unter anderem schlanke Dialoge ermöglichen und dem User die Möglichkeit bieten, ungewollte Dialogpfade schnell wieder zu verlassen, ohne den gesamten Dialog von vorne starten zu müssen.

Als zusätzliche Anforderungen erschienen Datensicherheit [33] und Kontextsensitivität wesentlich. Durch die Kontextsensitivität ist der Chatbot in der Lage, den Kontext des Users in die Gesprächsführung mit einzubeziehen und sich dadurch z. B. auf den Ticketverlauf zu beziehen [31].

Die Verwendung der Kategorien Kommunikationsfähigkeit und Problemlösungsfähigkeit entspricht den beiden Hauptanliegen der Kunden eines Servicedesks:

- *Kommunikationsfähigkeit.* Die inhaltliche Qualität des Chatbot-Outputs zeigt sich in einem hohen Anteil an richtigen und passenden und zudem freundlichen und empathischen Antworten. Dem Bereich der Kommunikation werden auch die allgemeine Usability sowie Sicherheitsaspekte zugeordnet. Die Effizienz des Chatbot-Outputs zeigt sich in kurzen Antwortzeiten und einer möglichst geringe Anzahl an Dialogschritten.
- *Problemlösungsfähigkeit.* Zu dieser Kategorie zählen die Anforderungen, mit denen die Bandbreite und Tiefe der Problemlösungsfähigkeit des Chatbots bewertet werden. Sie repräsentieren das Spektrum an Kundenanliegen in der Domäne eines Servicedesks und bewerten die Fähigkeit des Chatbots, ein Kundenanliegen abschließend zu lösen, bzw. einen als Mehrwert empfundenen Lösungsgrad zu erreichen.
- *Domänenspezifisches Kundenverständnis.* Im Schnittbereich zwischen Kommunikation und Problemlösung ist das Verständnis der Kundenanliegen angesiedelt, die der Chatbot verarbeiten kann. Zum einen ermöglicht das Verständnis von Fachbegriffen und typischen Anliegen eine reibungslose Kommunikation, zum anderen ist dieses Verständnis die Grundvoraussetzung für eine Lösung des Anliegens.

2.2 Konzeption des Bewertungsvorgangs

Die Bewertung, inwieweit der Chatbot die Anforderungen erfüllt, erfolgt in der hier vorgestellten Bewertungsmethode anhand seiner Antworten. Zur Datenerhebung wurde ein standardisierter Fragebogen entwickelt, dessen Fragen unabhängig von den Antworten immer gleich gestellt werden. Zu jeder Frage gehört eine erwartete Antwort, die bestimmte Qualitätskriterien erfüllen soll. Die Kriterien für die Bewertung der Antwortqualität wurden in Anlehnung an die Kriterien der Loebner-Preis-Evaluation festgelegt [20]. Es werden für jedes Kriterium abhängig von der Güte der Antwort null, ein oder zwei Punkte vergeben. Die Kriterien wurden für unsere Zwecke wie folgt angepasst:

Relevanz. Relevanz meint die Erkennung der Themen, die dem Kundenanliegen zugrunde liegen (z. B. Vertragskündigung). Für die Vergabe von zwei Punkten wird von der Antwort erwartet, dass das Thema bzw. die Absicht des Kunden durch den

Chatbot vollständig erkannt wurde. Entsprechend wird ein Punkt vergeben, wenn das Thema nur teilweise erkannt wurde und null Punkte, wenn der Chatbot das Anliegen nicht erkannt hat.

Korrektheit. Korrektheit meint die inhaltliche oder logische Korrektheit einer Antwort. Bei Fragen, die eine Vielzahl von Antworten zulassen, meint Korrektheit die inhaltliche Passung zu der zugehörigen Frage. Durch die Korrektheit einer Antwort kann die Logikfähigkeit, das Gedächtnis und der Umgang mit unerwartetem Input getestet werden. Die Punktevergabe erfolgt analog zur Relevanz und bewertet das Maß, in dem die Frage richtig bzw. passend beantwortet wurde.

Lösungsgrad. Der Lösungsgrad sagt aus, in welchem Maß das Anliegen des Kunden gelöst wurde (vollständig, teilweise, gar nicht). Dies kann je nach Anliegen durch das Geben von Informationen, das Anstoßen von Transaktionen, eine Problemdiagnose oder die vollständige Problembehebung geschehen. Für die Vergabe von zwei Punkten wird von der Antwort zudem erwartet, dass sie vollständig autonom durch den Chatbot gegeben wurde (ohne Weiterleitung an menschlichen Agenten). Das Kriterium Lösungsgrad erfordert zu jeder Frage die Definition der erwarteten Lösung.

2.3 Auswahl der Anforderungen und Ableitung der Fragen

Zur Durchführung der Bewertung wurde ein Bewertungsbogen konzipiert. Für die Erstellung des Bewertungsbogens wurden jene Anforderungen ausgewählt, die mittels eines standardisierten Fragebogens überprüft werden können (siehe Tabelle 1). Weitergehende Anforderungen, für deren Test ein Dialogverlauf nötig wäre, wie beispielsweise die Anforderungen nach einer möglichst geringen Anzahl an Dialogschritten oder nach einer dialogbasierten Störungsdiagnose, sind für diese Art der Überprüfung nicht geeignet und wurden daher ausgeschlossen.

Zu jeder Anforderung wurde eine oder mehrere Fragen bzw. Aussagen abgeleitet. Die ausgesuchten Fragen repräsentieren Anliegen aus dem Bereich des Servicedesks. Gleichzeitig wurde auf eine möglichst branchenunabhängige Auswahl der Fragen geachtet. Durch branchenunabhängige Fragen kann die Methode auf Chatbots im Servicedesk verschiedener Branchen angewandt werden.

Mehrere Fragen wurden insbesondere aus Anforderungen mit einem weiten Spektrum möglicher konkreter Kundenanliegen abgeleitet. Ein Beispiel dafür ist die Anforderung „beantwortet Informationsfragen" der Kategorie Problemlösungsfähigkeit, aus der Fragen zu Produkten und zur Rechnung abgeleitet wurden. Unter anderem wurde auf diese Weise die Frage „Welches ist euer günstigstes Produkt?" formuliert. Die Betreuung von Produkten und Services ist eine Kernaufgabe des Servicedesks. Gleichzeitig lässt die Frage offen, welcher Art die Produkte sind, so dass sie branchenunabhängig gestellt werden kann. Die Ableitung der erwarteten Antworten zeigt die Übersicht in Tabelle 2.

Tabelle 1. Durch standardisierten Fragebogen testbare Anforderungen

1 Kommunikationsfähigkeit
1.1 Zugänglichkeit
• gewährleistet die Erreichbarkeit der IT-Organisation [26] • kann als einzige Kontaktadresse funktionieren [16], [20] • beherrscht mehr als eine Sprache [28]
1.2 Kundenbehandlung
• erhöht Kundenbindung durch persönliche Note [8], [24] • kommuniziert freundlich und wertschätzend (abgeleitet aus [26]) • kann auf die Stimmungen der Kunden angemessen reagieren (abgeleitet aus [26]), [28] • kann mit Beleidigungen umgehen [8] • erinnert sich an vorangegangene Äußerungen [9]
1.3 Usability
• ist robust gegenüber Tippfehlern und erkennt Synonyme (bzgl. alltäglicher Begriffe) [8] • ermöglicht Abkürzungen für versierte Kunden [30] • bietet Conversation Listening [28] • nutzt Links für weiterführende Informationen [8], [22] • nutzt vorstrukturierte Antworten zur Steuerung der Nutzererfahrung [30] • bietet leichten Zugang zur Bedienungshilfe [8], [24]
1.4 Sicherheit und Kontext-Sensitiv
• gewährleistet Vertraulichkeit der Daten [33] • ist kontextsensitiv [28], [25]
2 Domänenspezifisches Kundenverständnis
• kann auf die Absichten der Kunden angemessen reagieren (abgeleitet aus [26]) • versteht Begriffe der Kundendomäne (Fachsprache der Kunden) (abgeleitet aus [26]) • robust ggü. Tippfehlern, erkennt Synonyme (bzgl. der domänenspezifischen Begriffe) [8] • nimmt entgegen: Informations-, Beratungs-, Produktwunsch, Anmerkung, Kompliment, Beschwerde, Störungsmeldung, Nachfrage zu bestehender Störung [16], [19] • kommuniziert verständlich [30]
3 Problemlösungsfähigkeit
3.1 Problemlösung
• beantwortet Informationsanfragen [16] • gibt Informationen und Vorschläge bei Beratungswunsch (abgeleitet aus [16]) • löst bei Produktwunsch Bestellungen aus [20] • nimmt Beschwerde entgegen und leitet Behebung oder Eskalation ein [16] • verbindet im Ausnahmefall mit einem persönlichen Ansprechpartner [17] • findet Fehlerursache (durch Messungen) heraus [19] • macht Lösungsvorschläge [19]
3.2 Ticketbearbeitung
• eröffnet ein Ticket (registriert Ticket) [16], [19], [20] • prüft Störungsmeldungen auf Service Level (Priorisierung) [16], [19], [20] • leitet Ticket an die richtige Stelle weiter und stößt weitere Bearbeitung an [16], [19] • verfolgt Ticket nach [16], [19] • koordiniert nachfolgende Support-Einheiten [16]
3.3 Integration (Anwendungen, Webseiten, Services)
• kann relevante Information auslesen und anbieten (abgeleitet aus [26]), [28] • kann notwendige Transaktionen anstoßen (abgeleitet aus [26]), [28]

Tabelle 2. Beispiele der Anforderungen an eine Frage

Kriterium	*Anforderung/ erwartete Antwort*
Relevanz	Der Chatbot versteht, dass es um das Produktangebot geht.
	Der Chatbot versteht, dass es um das günstigste Produkt geht. (2 Punkte)
Korrektheit	Es wird das günstigste Produkt genannt oder die Antwort geht auf das günstigste Produkt ein. (2 Punkte)
Lösungsgrad (mit Beispielen für Definition der erwarteten Lösung)	
2 Punkte	Kunde weiß, welches Produkt das günstigste ist. Er bekommt die Information direkt als Antwort oder als Link auf die entsprechende Seite.
1 Punkt	Kunde erhält einen Link oder Hinweis zu Produktseiten, muss dort aber selber nach dem günstigsten Produkt suchen. Oder Kunde wird mit einem Mitarbeiter verbunden und erhält so weitergehende Unterstützung.
0 Punkte	Der Kunde erhält keine Informationen dazu, wie er die günstigsten Produkte in Erfahrung bringen kann. Der Kunde wird lediglich an die allgemeine Telefonnummer oder die Webseite verwiesen.

3 Anwendung der Bewertungsmethode

Die Anwendung der Bewertungsmethode erfolgt durch eine „Befragung“ des zu bewertenden Chatbots sowie anschließende Bewertung der gegebenen Antworten anhand der spezifischen Kriterien und wird nachfolgend näher erläutert.

Der Fragenkatalog und die Bewertungskriterien bilden die zwei Grundelemente der Bewertungsmethode. Der Fragenkatalog enthält die aus den domänenspezifischen Anforderungen abgeleiteten Fragen bzw. Äußerungen. Durch die Eingabe dieser Fragen werden die Kommunikationsfähigkeit, das domänenspezifische Kundenverständnis und die Problemlösungsfähigkeit des Chatbots getestet. Die Bewertungskriterien sind das Mittel zur Beurteilung der Antwortqualität. Hier werden das Verständnis des Inputs und die Korrektheit bzw. Passgenauigkeit des Outputs sowie das Maß der durch die Antwort erreichten Problemlösung bewertet. Für die Durchführung der Bewertung wurde ein strukturierter Bewertungsbogen entwickelt, dessen Aufbau am Beispiel des Ausschnitts in Tabelle 3 erläutert wird.

Die Tabelle zeigt alle relevanten Elemente des Fragebogens. Auf der linken Seite sind die Fragen in ihren jeweiligen Kategorien und Unterkategorien aufgeführt. In der Spalte daneben steht die Anforderung an einen Chatbot, welche mit der Frage (teilweise auch mehrere) geprüft werden soll. Auf der rechten Seite wird der Chatbot mit seinen Antworten dokumentiert. Dabei wird nach Vor- und Nachteilen des Chatbots unterschieden. In die Felder können allgemeine Feststellungen, die bei der Durchführung der Methode auffallen, eingetragen werden. Darunter werden die Antworten des Chatbots mit zugehörigen Beobachtungen (Beob.) eingetragen, die aus den Antworten allein nicht hervorgehen. Daneben folgt die Bewertung der Kriterien Relevanz (R), Korrektheit (K) und Lösungsgrad (L). Die Summe (S) der Einzelbewertungen wird am Ende berechnet und eingetragen.

Tabelle 3. Ausschnitt des Bewertungsbogens

Kategorie: Problemlösungsfähigkeit *Unterkategorie:* Problemlösung		*Vorteile:*	*Nachteile:*				
Frage	*Anforderung*	*Antwort*	*Beob.*	*R*	*K*	*L*	*S*
Kannst du mich zu euren Angeboten beraten?	Beratungs-wunsch			0	0	0	0
Welches ist euer günstigstes Produkt?	Informations-wunsch			0	0	0	0
Ich möchte mich über dich beschweren.	Umgang mit Beschwerden			0	0	0	0

R = Relevanz, K = Korrektheit, L = Lösungsgrad, S = Summe

4 Evaluation der Bewertungsmethode

Zur Evaluation im Rahmen des Design Science Research Prozesses wurde eine Testreihe durchgeführt und dabei die Bewertungsmethode auf mehrere Chatbots angewandt. Im Hinblick auf den angestrebten Einsatzzweck der Methode wurde zur Untersuchung von Bewertungsqualität und Zweckmäßigkeit eine Reihe von Evaluationskriterien festgelegt: a) Objektivität, Validität und Reliabilität dienten als Kriterien der Bewertungsqualität sowie b) Domänenbezug, Branchenunabhängigkeit, Technologieunabhängigkeit, Usability sowie die innere Ordnung der Fragen als Kriterien der Zweckmäßigkeit.

Um diese Kriterien zu überprüfen, wurden fünf geeignete Chatbots für die Durchführung ausgewählt: Lisa (O_2), Julia (Vodafone Kabel Deutschland), Digitaler Serviceassistent (Deutsche Telekom), Chatbock (Klarmobil) und Clara (Otto). Gründe für die Auswahl waren die Vergleichbarkeit der Funktionalitäten der Chatbots innerhalb einer Branche (Telekommunikation mit Lisa, Julia, Chatbock und Digitaler Serviceassistent) und zwischen zwei verschiedenen Branchen (Telekommunikation und Versandhandel mit Clara) zu überprüfen. Ebenso kann durch diese Auswahl die Branchenunabhängigkeit des Fragenkatalogs geprüft werden. Durch die Auswahl von Chatbock, der im Unterschied zu den direkt auf der Unternehmenswebsite integrierten Chatbots auf dem Facebook Messenger aufbaut, kann außerdem die Technologieunabhängigkeit des Fragebogens untersucht werden.

4.1 Ablauf der Evaluation

Bei der Durchführung der Evaluation wurde jedem Chatbot die Frage: „Welches ist euer günstigstes Produkt?“ gestellt. Diese Frage adressiert die Anforderung der Bearbeitung von Informationswünschen. Für jede Frage des Fragebogens wurde das im folgenden beschriebene Vorgehen durchgeführt:

Als erstes wurde dem Chatbot die Frage gestellt. Die Antwort des Chatbots wurde dann in das Protokoll eingetragen. Diese lautete bspw. beim Chatbot Julia: „Links zeige ich Ihnen unsere günstigsten Kabeltarife“ ergänzt durch eine Auflistung von vier Tarifen inkl. Hyperlink. In die Spalte Beobachtung wurde nun eingetragen, dass auf der linken Seite Links zu den jeweils günstigsten Tarifen angezeigt werden. Danach konnte mit der Bewertung begonnen werden.

Der Chatbot hat erkannt, dass sowohl Informationen zu Produkten (hier Tarife) als auch von diesen Produkten die günstigsten gewünscht waren. Die Bewertung der Relevanz wurde daher mit zwei Punkten vorgenommen. Nach Prüfung der Links konnte bestätigt werden, dass diese Tarife die jeweils günstigsten der jeweiligen Kategorie sind. Die Korrektheit wurde deshalb mit zwei Punkten bewertet. Für die Bewertung des Lösungsgrads wurde die in Abschnitt 3.4 beschriebene Anforderung herangezogen. Da der Nutzer durch die Antwort weiß, welches die günstigsten Produkte sind und direkt Links zu diesen vorfindet, wurde auch der Lösungsgrad mit zwei Punkte bewertet. Allgemeine Feststellungen während der Evaluation konnten in den Zellen Vor- und Nachteile dokumentiert werden. Das vollständig ausgefüllte Protokoll für die Frage „Welches ist euer günstigstes Produkt?” ist in Tabelle 4 dargestellt.

Tabelle 4. Evaluationsdurchführung

Kategorie: Problemlösungsfähigkeit *Unterkategorie:* Problemlösung		*Vorteile:* schnelle Reaktionszeiten	*Nachteile:*				
Frage	*Anforderung*	*Antwort*	*Beob.*	*R*	*K*	*L*	*S*
Welches ist euer günstigstes Produkt?	Informations-wunsch	Links zeige ich Ihnen unsere günstigsten Kabel-Tarife.	Links zu den Tarifen	2	2	2	6

Dieses Vorgehen wurde für jeden der Chatbots und jede Frage wiederholt. Dabei wurden die Chatbots nacheinander bewertet. Aus den Bewertungen der Chatbots und den Beobachtungen während der Durchführung wurden die im folgenden Kapitel dargestellten Evaluationsergebnisse abgeleitet.

4.2 Evaluationsergebnisse

Die Evaluation der Bewertungsmethode hat verschiedene Stärken und Schwächen der entwickelten Bewertungsmethode aufgezeigt. Diese werden nachfolgend anhand der Evaluationskriterien für Bewertungsqualität und Zweckmäßigkeit diskutiert.

Objektivität, Validität und Reliabilität. Die Objektivität der Methode ist dann sichergestellt, wenn diese unabhängig von äußeren Einflüssen durchgeführt werden kann. Die Methode ist in ihrer aktuellen Form nicht vollständig objektiv. Das liegt vor allem daran, dass die erwarteten Antworten in einigen Fällen nicht vorher definiert werden können, da nicht klar ist, welche Fähigkeiten die Chatbots im Detail haben. Die Güte der Antwort kann in diesen Fällen nicht eindeutig bestimmt werden, so dass die Bewertung vom subjektiven Eindruck des jeweiligen Prüfers abhängt.

Reliabilität bedeutet, dass eine Messung wiederholt werden kann und sich die Ergebnisse nicht oder erklärbar unterscheiden. Die Reliabilität der Methode ist gegeben, da es sich um statische Fragen handelt und der Chatbot in der Regel bei gleichen Fragen die gleichen Antworten gibt. Eine Ausnahme hiervon stellt die Weiterentwicklung des Chatbots dar. Diese kann aber anhand der Differenz zur vorherigen Messung bewertet werden.

Eine Bewertung ist valide, wenn sie tatsächlich das misst, was mit ihr gemessen werden soll. Einfache Anforderungen, wie die in Abschnitt 4.1 angeführte Frage, können valide getestet und bewertet werden. Die Bewertung der Kriterien bei diesen Fragen mit null bis zwei Punkten hat sich bewährt, da sie nicht nur Ja-/Nein-Bewertungen zulässt, sondern eine differenzierte Bewertung ermöglicht. Die geringe Anzahl an Bewertungsalternativen erleichtert dabei die eindeutige Bewertung.

Eine Bewertung komplexer Themen, die von einem Chatbot nur im Dialog gelöst werden können, kann mit der aktuellen Bewertungsmethode nicht durchgeführt werden, was die Validität der Methode einschränkt.

Für einzelne Anforderungen kann durch den Bewertungsbogen nur die Text-/ Sprachausgabe des Chatbots bewertet werden. Beispielsweise ist in Bezug auf die Vertraulichkeit der Daten der tatsächliche Umgang des Chatbots mit Personendaten nicht über die Benutzerschnittstelle feststellbar. Bei der Weiterentwicklung der Bewertungsmethode wäre demnach zu prüfen, ob Anforderungen mit einem sehr breiten Spektrum möglicher Kundenanliegen, wie z.B. die Anforderung „beantwortet Informationsanfragen“, weiter detailliert werden können, um Unschärfen in der Bewertung zu reduzieren.

Zusätzlich ist der Fakt zu beachten, dass sowohl Relevanz als auch Korrektheit Einfluss auf den Lösungsgrad haben. Versteht ein Chatbot die Frage nicht oder ist nicht in der Lage, eine korrekte Antwort zu liefern, wird er in der Regel weder bei der Relevanz bzw. Korrektheit noch beim Lösungsgrad Punkte erhalten.

Allgemeine Feststellungen, wie die Reaktionszeit der Chatbots, werden zwar in den Vor- und Nachteilen betrachtet, aber nicht bewertet. Dabei können gerade die hier festgestellten Eigenschaften des Chatbots erhebliche Auswirkungen auf den Kundennutzen haben. Um die Methode in dieser Hinsicht zu verbessern, können Erkenntnisse von Kuligowska [8] für die punktemäßige Bewertung von beobachtbaren Merkmalen einbezogen werden.

Domänenbezug. Der Domänenbezug konnte mit der Bewertungsmethode in der aktuellen Form nur teilweise erreicht werden, da typische Kundenanliegen im Servicedesk eine Dialogführung für ihre Erfassung und Bearbeitung benötigen.

Branchenunabhängigkeit. Die Branchenunabhängigkeit der Fragen konnte festgestellt werden. Dieses wurde durch die Auswahl von allgemeinen Fragen oder den Einsatz von Variablen (Bsp.: „Ich möchte mich über <Unternehmen> beschweren.”) in den Fragen sichergestellt. Chatbots aus zwei verschiedenen Branchen können dadurch in Bezug auf ihren Nutzen hinsichtlich typischer Servicedesk-Themen miteinander verglichen werden. Auch innerhalb einer Branche bietet die Methode Vergleichsmöglichkeiten.

Technologieunabhängigkeit. Mit Fokus auf die Technologieunabhängigkeit zeigte sich der Fragebogen in der Testreihe als ausreichend abstrakt, da die Nutzbarkeit des Fragebogens nicht durch die zugrundeliegende Technologie beschränkt wird. Aus Kundenperspektive muss der Nutzen unabhängig von der eingesetzten Technologie eintreten. Ergebnisse zu technologiebedingten

Nutzungsbeschränkungen des Chatbots können im Kopfteil des Fragebogens vermerkt werden (Bsp.: „Facebook-Account notwendig“ für den Chatbock).

Usability. Aus dem Blickwinkel der Usability zeigte sich, dass die Anwendung der Methode vergleichsweise wenig aufwendig ist. Für die Bewertung eines Chatbots wurden knapp zwei Stunden aufgewendet, technische Hilfsmittel sind nicht erforderlich. In einer der folgenden Iterationen zur Verbesserung der Methode soll dennoch versucht werden, die Anwendungszeit weiter zu verringern, etwa durch (Teil-)
Automatisierung.

Innere Ordnung. Die innere Ordnung der Fragen konnte durch die Zuordnung zu den Kategorien und Unterkategorien gewährleistet werden.

5 Fazit und Ausblick

In Beantwortung der Forschungsfrage ist eine Bewertungsmethode entstanden, die ausgehend vom fragebogenbasierten Vorgehen der Loebner-Preis-Methode die Spezifika eines im Servicedesk agierenden Chatbots integriert und dadurch die Nutzenbewertung eines solchen Chatbots durch den konstruierten Fragebogen mit zugehöriger Bewertungsmethode domänenspezifisch unterstützt.

Durch die systematische Erfassung der vielschichtigen Anforderungen an einen Chatbot als Grundlage zur Erstellung des Bewertungsbogens zeigt die Arbeit außerdem die Mannigfaltigkeit der von einem Servicedesk-Chatbot abzudeckenden Themenbereiche auf. Darüber hinaus konnte durch die Evaluation der Bewertungsmethode anhand von fünf Chatbots eine praxisorientierte Nutzungsmöglichkeit des erstellen Artefaktes aufgezeigt werden, die sich durch die Anwendbarkeit in verschiedenen Branchen sowie Unabhängigkeit von der technischen Implementierung auszeichnet.

Die aufgezeigten Beschränkungen des bisherigen Methodenentwurfs motivieren weiterführende Forschungsarbeiten mit dem Ziel, die Bewertungsmethode bspw. durch erweiterte Evaluation mit einer größeren Anzahl von Chatbots zu verbessern. Vorrangig sollte nach Möglichkeiten der Integration von Dialogverläufen in die Bewertung gesucht werden. Weiterhin sind Automatisierungspotenziale bei der Anwendung der Bewertungsmethode erkennbar, die zu einer Reduzierung des manuellen Aufwands führen könnten. Als reizvoller Ansatz wäre hier die Entwicklung eines Chatbots zur automatisierten Befragung des zu bewertenden Chatbots vorstellbar. Schließlich könnte geprüft werden, ob sich die Methode auch für eine reifegradbasierte Untersuchung von Chatbots erweitern ließe.

Literatur

1. Prifti, L., Heininger, R., Hecht, S., Krcmar, H.: Service Desk Requirements for Cloud Service Providers. In: Kundisch, D., Suhl, L. Beckmann, L. (eds.) Tagungsband Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014), pp. 1173-1184. Universität Paderborn (2014)

2. Knapp, D.: A Guide to Customer Service Skills for the Service Desk Professional, 3rd ed. Course Technology, Boston (2011)
3. Schatsky, D., Gratzke, P.: The conversational enterprise – how chatbots can answer questions for both customers and employees. Deloitte University Press (2016)
4. Heo, M., Lee, K.J.: Chatbot as a New Business Communication Tool. The Case of Naver TalkTalk. Bus. Commun. Res. Pract. 1, 41–45 (2018)
5. Fittkau & Maaß Consulting: W3B Exklusivstudie Chatbots und Social-Bots., https://www.developer-week.de/studie-zu-chat-und-social-bots/ (Accessed on 22.07.2018)
6. Stoeckli, E., Uebernickel, F., Brenner, W.: Exploring Affordances of Slack Integrations and Their Actualization Within Enterprises – Towards an Understanding of How Chatbots Create Value. In: 51st Hawaii International Conference on System Sciences, DOI: 10.24251/HICSS.2018.255 (2018)
7. Braun, D., Hernandez Mendez, A., Matthes, F., Langen, M.: Evaluating Natural Language Understanding Services for Conversational Question Answering Systems. In: SIGDIAL 2017 Conference, pp. 174–185. Association for Computational Linguistics, Saarbrücken, Germany (2017)
8. Kuligowska, K.: Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents. Professionals Center for Business Research 2, 1–16 (2015)
9. Shawar, B.A., Atwell, E.: Different measurements metrics to evaluate a chatbot system. In: Weng, F., Wang, Y.-Y., Tur, G., Hu, J. (eds.) Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, NAACL-HLT- Dialog 2007, pp. 89–96. Association for Computational Linguistics, Stroudsburg, PA, USA (2007)
10. Stieglitz, S., Brachten, F., Kissmer, T.: Defining Bots in an Enterprise Context. In: 39th International Conference on Information Systems, San Francisco (2018), angenommen, https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1168&context=icis2018 (Accessed on 13.11.2018)
11. Brandtzæg, P.B., Følstad, A.: Why people use chatbots. In: Kompatsiaris, Y., Cave, J., Satsiou, A., Carle, G., Passani, A., Kontopoulos, E., Diplaris, S., McMillan, D. (eds.) Internet Science. 4th Int. Conf., INSCI 2017, Thessaloniki, Greece, pp. 377–392. Springer International Publishing, Cham (2017)
12. Jafarpour, S., Burges, C.J.C.: Filter, Rank, and Transfer the Knowledge: Learning to Chat, Microsoft Research Technical Report MSR-TR-2010-93 (2010)
13. Xu, A., Liu, Z., Guo, Y., Sinha, V., Akkiraju, R.: A New Chatbot for Customer Service on Social Media. In: Mark, G., Fussell, S., Lampe, C., Schraefel, M.C., Hourcade, J.P., Appert, C., Wigdor, D. (eds.) Conference on Human Factors in Computing Systems, CHI '17, pp. 3506–3510. ACM Press, New York, NY, USA (2017)
14. Følstad, A., Brandtzaeg, P.B., Feltwell, T., Law, E.L.-C., Tscheligi, M., Luger, E.A.: Chatbots for Social Good. In: Mandryk, R., Hancock, M. (eds.) Extended abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–4 (2018)
15. Jäntti, M., Shrestha, A., Cater-Steel, A.: Towards an Improved IT Service Desk System and Processes: A Case Study. Int. J. on Advances in Systems and Measurements 5(3&4), 203–215 (2012)
16. Kopperger, D., Kunsmann, J., Weisbecker, A.: IT-Servicemanagement. In: Tiemeyer, E. (ed.) Handbuch IT-Management – Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis., pp. 162–179. Carl Hanser Verlag, München (2013)
17. Fujitsu Services Ltd.: Der Service Desk der nächsten Generation. Unsere Welt steckt mitten in einer digitalen Revolution.,

https://sp.ts.fujitsu.com/dmsp/Publications/public/white-book-service-desk-digital-exec-sum-de.pdf (Accessed on 17.06.2018)
18. Beims, M.: Das ITSM der Zukunft. Wie Digitalisierung das ITSM verändert, https://www.cio.de/a/das-itsm-der-zukunft,3578603 (Accessed on 17.04.2018)
19. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77 (2007)
20. The Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB): Loebner Prize 2014 Contest Archive, Loebner Prize Selection Process, http://www.aisb.org.uk/events/loebner-prize, 04.07.2018)
21. Mauldin, M.L.: Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. In: 12th National Conference on Artificial Intelligence (AAAI-94), pp. 16–21, Seattle (1994)
22. Legg, S., Hutter, M.: Tests of Machine Intelligence. In: Lungarella, M. et al. (eds.): 50 Years of Artificial Intelligence. Lecture Notes in Computer Science, vol 4850 pp. 232–242. Springer, Berlin, Heidelberg (2007)
23. Vasconcelos, M., Candello, H., Pinhanez, C., Donizetti dos Santos, T.: Bottester: Testing Conversational Systems with Simulated Users. In Brazilian Symposium on Human Factors in Computing Systems, IHC '17 (2017)
24. Walker, M.A., Litman, D.J., Kamm, C.A., Abella, A.: PARADISE. In: Cohen, P.R., Wahlster, W. (eds.) 35th Annual Meeting on Association for Computational Linguistics, pp. 271–280. Association for Computational Linguistics, Morristown, NJ, USA (1997)
25. Proehl, T., Erek, K., Limbach, F., Zarnekow, R.: Topics and Applied Theories in IT Service Management. In: 46th Hawaii International Conference on System Sciences, DOI: 10.1109/HICSS.2013.555 (2013)
26. Elsässer, W.: ITIL einführen und umsetzen. Leitfaden für effizientes IT-Management durch Prozessorientierung. Hanser, München (2005)
27. Fischlin, R.: Leitfaden ITIL-Service Desk. In: Müller, P. (ed.) 20. DFN-Jahrestagung, DFN 2006, pp. 95–104 (2006)
28. Smiers, L.: How can Chatbots meet expectations? Introducing the Bot Maturity Model, https://www.capgemini.com/2017/04/how-can-chatbots-meet-expectations-introducing-the-bot-maturity/ (Accessed on 17.04.2018)
29. Nielsen, J.: 10 Usability Heuristics for User Interface Design, https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed on 17.04.2018)
30. Scott, K.: Usability Heuristics for Bots, https://chatbotsmagazine.com/usability-heuristics-for-bots-7075132d2c92 (Accessed on 17.04.2018)
31. Makadia, M.: Challenges of Building an Intelligent Chatbot, https://dzone.com/articles/challenges-in-building-an-intelligent-chatbot (Accessed on 23.07.2018)
32. Härtl, M.: Konzeption und Realisierung der technischen Unterstützung eines zentralen IT-Service-Desk mit OTRS an der T[echnischen] U[niversität] M[ünchen], Diplomarbeit (2007), http://www.nm.ifi.lmu.de/pub/ Diplomarbeiten/haer07/PDF-Version/haer07.pdf (Accessed on 17.04.2018)
33. Conrad, C.: Datenschutz bei neuen Kommunikationskanälen wie Chatbots und Messenger-Diensten: Der automatisierte Kundenkontakt der Zukunft, https://www.datenschutz-notizen.de/datenschutz-bei-neuen-kommunikationskanaelen-wie-chatbots-und-messenger-diensten-der-automatisierte-kundenkontakt-der-zukunft-3520019/ (Accessed on 17.04.2018)

Understanding the Benefits of Agile Software Development in Regulated Environments

Jens Karrenbauer, Manuel Wiesche, Helmut Krcmar

Technische Universität München, Chair for Information Systems, Munich, Germany
jens.karrenbauer@tum.de, {wiesche, krcmar}@in.tum.de

Abstract. Agile software development has become increasingly popular in recent years. Applying agile methods, companies expect flexible planning, early delivery of the software product, and a continuous improvement of the development process itself. However, in regulated environments the use of agile development is not yet common practice. In such environments, various regulatory requirements apply which affect the software development process. This paper examines the use of agile software development in the regulated medical device industry and explores reasons for using agile methods although their use is limited. We interviewed agile software development teams in three different companies using semi-structured interviews. Using grounded theory methodology, we identify reasons why companies are using agile methods, even though problems and barriers exist. Our main achievement is the development of four categories, which describe the benefit of agile software development in regulated environments. These categories are master complexity, reduce effort, improve usability, and promote collaboration.

Keywords: agile software development, regulated environments, benefit of agile software development, medical device industry

1 Introduction

For some time, companies adopted agile methods, such as Scrum, within their software development process. Reasons for agile adoption are benefits such as flexible planning, early delivery of the software production, or a continuous improvement of the software development, which results in better quality software [1]. However, companies often find it difficult to implement agile methods due to challenges in the organizational, cultural, and human context [2]. In regulated environments, this is intensified by various regulatory requirements that must be met, while at the same time reacting flexibly to market or customer requirements is necessary [2].

One characteristic of regulated environments is the multitude of compliance procedures, regulations and standards that have to be considered in the software development process. There are various organizations and associations, such as the International Organization for Standardization (ISO) or the Food and Drug Administration (FDA),

14^{th} International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

which formulate specific requirements [3]. Furthermore, self-imposed requirements apply which are not mandatory for an industry, but which are generally recognized standards required by customers [4].

One example is the medical device industry, which is highly regulated. Companies that develop software for the US market must guarantee that their software complies with all FDA requirements.[1] For all stakeholders of these organizations, it is important to see that the system developed was designed safely and effectively for the intended purpose [6]. Breaching these requirements can quickly cause economic damage to a company [7]. For instance, a violation of regulatory requirements can lead to sanctions such as product recalls, product seizures or export restrictions [8].

A company in the medical device industry must, therefore, meet the challenge of reconciling the both existing and changing, strict regulations with the agile approach and at the same time constantly reducing development cost [2]. Thus, understanding why agile methods are used in regulated environments despite these barriers is important. This paper attempts to answer the following research question: *What are typical benefits of the usage of agile software development in a regulated environment?* To explore this question, we conducted 20 interviews with members of agile software development teams in three different companies from the medical device sector. Using the Grounded Theory Methodology (GTM) for data analysis, our main achievement is the identification of four categories, which describe the benefit of agile software development in regulated environments.

The remainder of the paper is structured as follows. In the next section, we provide a brief introduction of the background literature. This is followed by an introduction of our research approach as well as a presentation of the relevant cases. We then present our results and discuss the findings in detail. The paper concludes by highlighting the contributions that our study makes to research and practice.

2 Background

2.1 Agile Software Development in Regulated Environments

Agile software development methods became more and more popular in recent years. Different methods have been classified as agile, e.g. eXtreme Programming (XP), Crystal, Feature Driven Design, or Scrum. Scrum is one of the most widely used agile development methods in practice [3]. This method has been adopted in many sectors of the software industry. Companies expect flexible planning, early delivery, and continuous improvement of the development process [1]. Agile methods promise to reduce development time, increase product quality, and reduce development costs [9].

[1] The EU directives for medical devices are in a period of change. On 25 May 2017, the new European Medical Devices Ordinance (MDR) came into force with a transition period of three years. This imposes extensive new requirements on the clinical evaluation of devices [5] and assimilates the requirements of the FDA and EU.

Agile was originally intended to be used by small teams in non-safety-critical projects and in co-located environments [10] [11]. However, agile methods need to be tailored to suit the needs of different circumstances [12]. Agile has been adopted to other contexts, including virtual environments [13], global software engineering [14], complex systems or Capability Maturity Model (CMM) environments [15] [16], as well as regulated environments [3].

Especially in regulated environments scholarly research finds puzzling results regarding the use of agile methods. The main conclusion of [3] is that agile practices and regulated environments are not necessarily incompatible. Others say that agile practices are unsuitable and can only be used successfully in combination with plan-based methods [17]. Most of the literature recommends an adaptation of agile approaches (e.g. [9], [18], [19]), while some consider it is too early altogether to apply agile methods in such environments (e.g. [15], [20]). One possible explanation for these inconsistent empirical findings are different regulated environments with different focuses.

Table 1: Relevant laws and regulations in the field of medical devices with their impact on the process.

Law/ Regulation	*Comment/Impact*
Focus: Control instrument for demonstrating the safety and medical-technical performance of medical devices.	
EU Medical Device Directive (MDD), EU Medical Device Regulation (MDR) *[23] [24]*	• Define software as a medical device. • Divide medical devices into four classes which are used as the basis for the requirements to be implemented and the scope of an inspection by the authorities. • Require state of the art validation of the software, considering the principles of the software life cycle, risk management, validation and verification.
Focus: Software Life Cycle Processes for Medical Device Software	
AANIS/AAMI/DIN EN/IEC 62304 *[25] [22]*	• Requires activities and tasks within the software life cycle required for the safe development and maintenance of medical device software. • Requests a prior determination of what is expected of the software and subsequent proof that the use of the software fulfils these intentions without causing unacceptable risks.
Focus: Quality management systems for medical device manufacturers.	
ISO 13485, 21 CFR 820 *[26]*, *[27] [21]*	• Requires the development of internal policies and procedures describing the QMS. • Requires the creation of Standard Operation Procedures (SOP). • Requires the creation and documentation of a comprehensible software development process. • Low-level processes must be consistent with those at a higher level and provide a coherent and consistent approach across the organisation that meets regulatory requirements.
Focus: Application of risk management to medical devices	
ISO 14971 *[28] [22]*	• Calls for the introduction of systematic management of the risks associated with the development and operation of a medical device. • Requires consideration of product risks for patients, operators, others, and the environment. • Requires continuous monitoring of risks, even after delivery of the product.

2.1 Requirements for Agile Software Development in Regulated Environments

For the medical device industry, different laws and regulations apply with regards to the software development process. Various organizations and associations such as the ISO and the FDA, formulate and publish regulatory requirements for product development [3]. For the development of software this implies a need for high quality, security and reliability. Table 1 illustrates the most relevant laws and regulations for the EU and the US and provides more information about potential impacts of these requirements on the software development process.

The table shows that the specific regulatory requirements differ between the countries. These international standards and guidelines make the medical device software development process unique [21]. Especially the IEC 62304, which places demands on life cycle processes for medical device software, is recognized by many regulatory bodies around the world as the *gold standard* [22].

There are strict requirements for the reliability and traceability of the products and their development process in the medical device industry. From these specific requirements, general rules can be derived which apply in a regulated environment in general. Table 2 highlights these common requirements. For instance, quality is one of the most important characteristics that a process and its software must show. Additionally, it is imperative to plan the entire project, as well as document all the activities done. Companies adopting agile methods face barriers when complying to the agile principles, for example, a lack of documentation or up-front planning. [29] Also, there are different agile barriers for the development teams which have an impact on the performance. Examples are the occurrence of subgroups or the relationship of personality models [30] [31] [32].

Table 2: Typical examples of requirements in a regulated environment [2]

Topic area	*Description*
Ensuring quality	planning, implementation and proof of quality assurance measures [6]
Ex-ante planning	additional process steps (e. g., risk analysis or additional tests) [15]
Roles & responsibilities	mandatory roles (e. g., security expert or a quality manager) [33]
Documentation	extensive, development-related documentation (e.g. requirements documentation or test specifications) [8]

The impact of these requirements leads to different problems, for instance during the overall process implementation, during the transition to the testing phase, or within the documentation. For example, all tasks and activities must be completed before a release takes place (including a complete documentation). This also means, software cannot be delivered without a complete documentation [18]. Additionally, it is not worth to deliver small releases in short cycles as the additional work to be done is extensive. Also, the costs of refactoring will become quickly very high as soon as a change involves a re-verification and re-evaluation of the artifacts concerned [20].

Literature showed several examples for such problems, which redundant occur in this context. That is the reason why we want to clarify and understand the motivation and benefits for companies in a regulated environment to use agile methods.

3 Research Approach

To understand why agile is used in regulated environments requires a rich understanding of the field and yields several potential explanations. We, therefore, find an inductive, qualitative research design appropriate. We conducted 20 interviews in three different cases and used the GTM for data analysis. Based on the case studies and the results of the GTM, four categories which describe the benefits of agile software development in a regulated environment have been elaborated.

Table 3: Background information of the interviewees

				experience (years) in			*agile*	
ID	*C*	*Role*	*Background*	*CO.*	*IN.*	*agile*	*exp.*	*Exp.*
1	A	Agile Coach	Computer Science	10	10	8	S, N	MD, AV
2	A	Scrum Master	Medical Technology	6	11	14	S	MD
3	A	Product Owner	Developer	16	3	10	S, K	MD, AT
4	A	Architect	Telecomm.	10	10	4	S	MD, AV
5	A	Program Mgr.	Medical Devices	18	18	18	S	MD
6	A	Verification	Telecomm.	13	6	4	S	MD, AV
7	A	Risk manager	Biology	17	17	4	S	MD
8	B	Req. Eng.	IT Specialist	12	12	6	S	MD
9	B	Ergonomics Mgr.	Medical Informatics	2	8	2	S	MD
10	B	Product Owner	Industrial Eng.	4	4	4	S, K	MD
11	B	Scrum Master	Industrial Eng.	10	17	6	S	MD
12	B	Architect	Medical Informatics	10	15	6	S	MD
13	B	Risk Manager	Mechatronics Eng.	1	16	N/A	N/A	MD
14	B	Tester	Computer Science	5	4	4	S	MD
15	B	Developer	Electrical Eng.	12	12	12	S	MD
16	B	Program Mgr.	Electrical Eng.	17	17	8	S	MD
17	C	Program Mgr.	Electrical Eng.	5	5	7	S, XP	MD, AT
18	C	Scrum Master	Medical Informatics	7	14	10	S	MD
19	C	Developer	Computer Science	7	7	10	S, XP	MD
20	C	Product Owner	Computer Science	6	10	6	S, K	MD, RW

Legend: C = case; CO = company; IN = industry; MD = medical devices; S = Scrum; K = Kanban; N = Nexus; XP = eXtreme Programming; Exp. = Experience with regulated environments; AV = aviation; AT= Automotive; RW = Railway

3.1 Research Method

For sampling our data, we chose a holistic multiple-case design, where we identified several distinct case organizations to be included in our sample [34]. We identified three companies, each of which is regarded as a separate case. A description of these cases is provided in the following.

In all cases semi-standardised interviews were conducted to collect data [35]. An interview guideline was developed which served as a basis for the interviews. This guideline included questions on known regulations, agile procedures used in the company, characteristics of the implemented active development process, and one's own opinion on existing conflicts. Additionally, we asked questions about the benefits and existing barriers of the usage of agile software development in the regulated environment. The aim was to interview each role of the agile software development team in order to gain a comprehensive picture of the benefits using agile software development methods in regulated environments [35].

We conducted 20 interviews. Each interview lasted between 40 - 90 minutes. Table 3 provides background information of all interviewees. After each interview, a theoretical memo was prepared to record our impressions and thoughts about the interviewee [36]. The interviews were transcribed and analysed with Atlas.ti.

For data evaluation and analysis we followed Grounded Theory Methodology (GTM) [37]. In the first step of the analysis, initial concepts were identified in the data. These were divided into categories (open coding). Concise and self-explanatory codes were used to create the categories. Next, the open codes were divided into groups by selective coding. Due to the large number of codes, we also worked with subcategories of the respective categories.

3.2 Cases

We identified three companies from the medical device industry which are already using agile methods within their software development process. The first case, INFUSE, is a software design house with experiences in the automotive, safety & security, aerospace, and medical device sectors. The project focuses on the development of software for an automatic infusion pump. The implementation of Scrum is close to the framework. Concurrent to the agile process, there is a risk management, a verification, and a configuration management process. Issues or requirements of the concurrent processes are given as requirements in the Scrum process. In our view, the company does not distinguish between regulatory requirements and product requirements as they want to use the opportunity of developing a high-quality software product.

The second case, XRAY, is a company in the dental industry. The interviews were based on the development of a software, which is used for the imaging of X-ray machines (image display, image enhancement). A special feature is that the software can be used as a variant with other imaging devices (e.g. industrial area). Currently, there are two agile Scrum teams working on this project. There are also many other roles involved, e.g. an Ergonomics Manager or a Requirements Engineer. At the start of a new project (or a new release) all requirements have already been defined. The whole software development process is very extensive and rigid. The interpretation of the existing regulations is always enforced strictly.

The third case, DATA, is a provider of products and solutions for ophthalmology, neurosurgery, dentistry and oncology. The software can be used to collect, process and archive data from their system as well as third-party devices. In addition, the data is prepared for the physician and the information is made available for diagnostics. The implementation of the Scrum framework at DATA is a pragmatic approach. Nevertheless, it takes them an average of one year until a new release takes place.

4 Results

4.1 Software Development Process in the Cases Applied

The three projects are similar at a high level. The same regulations apply to all three cases, since all software products are grouped in the same classification of medical devices (IIb). In addition, the same agile practices are predominantly used. In all cases the embedding of the software development process in a higher-level product development/medical product process took place and run as a V-model. For the software development they move into the agile process. After several iterations within Scrum, the agile process ends, and the V-model is completed with verification/system tests and other necessary activities. Table 4 provides an overview of the three cases.

Table 4: Comparison of the three cases

	Case A (INFUSE)	*Case B (XRAY)*	*Case C (DATA)*
Employees	190	1000	3000
Product	Software for an infusion pump	Software for the imaging of x-ray machines	Data Management System for Ophthalmology
Form of Software	Embedded	Stand-Alone	Stand-Alone
MD classification		IIb	
Runtime	- Started: 2012 - Scrum: Since 2014 - End: 2019	- Feasibility study: 2011 - Started: 2012 - Current version: 2.3	- Started 2008 - Current version: 4.2.1
Relevant regulations	ISO 13485, 14971, 60601, 62304, FDA	ISO 13485, 14971, 60601, 62304, FDA	ISO 13485, 14971, 60601, 62304, FDA
Team Size	Scrum-Team: 8 members (6 developer, 1 PO, 1 SM)	Scrum-Team: 12 members (10 developer, 1 SM, 1 PO)	Scrum Team: 4 members, 1/3 PO, 1/3 SM (PO & SM are shared with 3 other teams)
Used agile practices	Burn-Down-Chart, Code Ownership, Code Refactoring, Continuous Integration, Pair Programming, Mob Programming, Retrospective, User Stories,	Burn-Up-Chart, Code Refactoring, Continuous Integration, Pair Programming, Retrospective, User Stories,	Burn-Down Charts, Code Ownership, Code Refactoring, Continuous Integration, Pair Programming, Retrospectives, User Stories,
Agile strategies	- embedding agile in existing V-model based development - using agile as toolbox of different procedures - “mini-Vs” within each sprint		

In order to compare these cases, we need to take a deeper look into the processes and have, therefore, identified three main characteristics by using GTM which are handled differently. These characteristics are *implementation of requirements*, *documentation* and *final integration/system test*. Implementation of requirements means time and procedures for the implementation of requirements. Documentation describes the time and procedure of the documentation. Final integration/system test characterises the handling of these final tests. Table 5 compares the three cases.

Table 5: The software development process in the three different cases

	INFUSE	*XRAY*	*DATA*
Implementation of requirements	- user stories are partly defined in advance, or they are created dynamically parallel to the development	- strict modular processing of the various steps - requirements are defined and approved in advance	- development is based on the defined user story and the acceptance criteria - no formal specification in advance
Documentation	- documentation is usually made in parallel or after development	- without finalized documentation, the software may not be implemented.	- no current, formal documentation on the development is required - developers are free to decide how they want to document the development
Final Integration / System test	- a full system test is performed after the last sprint and before release	- a full system test is performed after the last sprint and before release	- there is no completely system test - final tests of the main operating functions take place

4.2 Benefits of Using Agile Methods in Regulated Environments

If one looks at the problems and barriers identified in the previous chapters, it is noticeable that these areas match the aspects that are considered as an incentive for the introduction of agile development methods. The coding of the data revealed four categories of motivation to introduce agile development methods. These four categories are *master complexity, reduce effort*, *improve usability*, and *promote collaboration*.

Master complexity

Software development and software itself have grown more and more complex in the previous decades. With new technological possibilities, increasingly complex systems found their way into medicine. In this context, agile methods help the software development team to deal with complexity-related issues such as compatibility with modules, integration with hardware, or handling all the regulations applying to the software development process.

In contrast to an agile approach, the classic V-model provides a complete definition of all requirements at the beginning. In the case of complex systems, this is sometimes difficult to implement, as the entire set of requirements must be described initially. If a V-model is used for software development today, it leads to additional efforts as existing developments have to be continuously changed. Regardless of the selected process model, an iterative procedure is required to master the complexity.

There are different possibilities how agile helps to master the complexity in a software development project. For instance, agile is suited very well if the definition of the final product is not quite clear at the beginning of the project: "*Actually, as soon as it becomes more complex and at the beginning is still a certain blurriness in the whole,* *then it is always helpful, even if it is only that I go through the first phases agile and iteratively." (INFUSE).* Additionally, iterative learning is nowadays almost indispensable for complex system development:

"*Mastering the complexity is in an agile team, where I don't have quite rigid requirements, and that has to be on the table in three years (...). But, I have a vision, and also the vision I adjust in my reflections constantly after a bit. Of course, I can't throw it all over the place, that there will be something completely different. Because then I produced for the trash can. But that it is just a matter of bringing a vision into the world and with the knowledge I learn about the iterations, that it really grinds itself so finely that in the end what I need really comes out.*" *(XRAY).*

Another point and what we already mentioned is, that complex software can no longer be defined completely in advance. The reason for this is the high uncertainty regarding the features to be developed, as well as the difficulty of understanding complex interdependencies in advance: "*I don't have the chance to initially describe all requirements, all interfaces in such a way that I can easily implement them.*" *(XRAY).* The following quote explains this in more detail:

"The whole world and technology and systems are becoming increasingly complex. This means that we are networked, all devices talk to each other, and the complexity is meanwhile on one level that it is not controllable in this classical process model topic, like also a V-model." (INFUSE).

The degree of complexity in such an environment became so high, that the traditional process models are no longer suitable: "*And that's why a big key of agility is that none of the very complex systems, no single person or individual, is able to describe the system in such a way that it is then processed by a team, and in the end what I really want, and need comes out.*" *(INFUSE).*

This illustrates how the use of agile methods helps software development teams to deal with complexity. In a regulated environment, an agile approach is relatively rigid, but still more dynamic than in a classical process model.

Reduce effort

An agile approach is far more efficient than a classic process model. It can help software development teams to reduce effort in a better way than classic process models do. Efficiency can be achieved by saving time (and thus increasing speed) or preventing waste (and thus reducing the number of tasks). Speed is an essential factor in software development. With agile methods and procedures, the development of the product begins earlier in the development process.

Generally, an agile approach is leaner than a classic approach and, thus, is better able to reduce effort in a software development process. In addition, a minimum viable product is usually available at an early stage: *"Because the agile is more like saying with the whole toolset, you start working much earlier on what the Minimal Viable Product is. That means you try to understand the problem much better, so you don't generate waste by specifying something that no one needs. And there is also such a bit of cultural change necessary." (INFUSE).*

Additionally, with an agile approach, it is possible to react and develop faster. This is a reason why companies are choosing this method: "*That was our main motivation* *to actually introduce agility. Namely to be able to react faster to changing market requirements in our process world as well, and in development." (DATA).* Also, traditional activities already planned are often not pursued further, resulting in unnecessary planning effort. In addition, requirements are often implemented in the software that are not required in the end. Software developers refer to this as "waste":

"[..] a motivation to originally introduce agility was that the predictability about what is the feature scope had nothing to do with what was then in the product requirement spec. So, there was something in it, but during the development period, it felt 50% completely different. And then it had made absolutely no sense to stay in this rigid process model now." (DATA).

"[...] and because he knows that, he defines what he thinks what will be needed in five years. The next big thing. And most of it is just over-specified. This means that you simply create a solution to problems that no one has." (DATA).

In summary, this means that in an agile environment only needed requirements are implemented: *"And then I have all the requirements that I want to implement, which I would not find with a classical approach. Because then I add insane things and have as requirements that no one really needs, and I would still forget a lot*." *(DATA)*. This, combined with regular feedback, leads to an efficient development process. Although mandatory activities for the provision of software must take place for a regulatory environment, these are, as expected, equally mandatory in a classical process model. The advantage in an agile environment is that the waste previously generated with conventional methods is no longer necessary, which reduces the effort and increases the efficiency in software development.

Improve usability

Even though some essential aspects of agile methods, such as early delivery or more flexible planning, are difficult to apply in a regulated environment, there are other facets that are beneficial and suitable for that environment. Software development teams can use these lean frameworks with a limited number of rules and specifications in various scenarios. Due to their simplicity, they are adaptable and can be combined with various agile and non-agile methods and procedures.

One example for the improvement of the usability is the fact that an agile approach leads to more innovation through employees. "*I can let the employees go, those who are normally motivated, they do it. And others, who are highly motivated, bring in more innovation and so on*." *(INFUSE).* Another example is the continuous process improvement which is very well supported by the Scrum framework. "*[...] the knowledge, and then contribute it back to the improvement process. So continuous process improvement. That's good." (XRAY)*. This is accompanied by regular feedback and adjustments which are carried out within the iterations. "*And what is definitely useful are the short iterations, where you quickly realize, ok, maybe we are on the wrong path after all, maybe we still have to readjust*." (*XRAY).*

Additionally, while using Scrum, it is possible to be nearer to the reality: "*And it's just the same, as always, the problems are actually always human. And that doesn't matter which process it is, you will always have that. Only the agile process helps us to be much more in reality than the waterfall process would do." (DATA)*. Last but not least, the documentation within a software development project is simpler than in a traditional process model: "*The documentation is easier to think through in the classic scrum, product backlog, product backlog item, test, idea, item, than in that (..), gives me a request and I turn you the request over, and then it is a test case." (INFUSE).*

Agile methods have several advantages which improve the usability within a software development process. Continuous process improvement is characteristic for an agile environment. This is particularly helpful in a regulated environment due to the existing rigidity and necessary additional activities. This includes providing regular feedback to stakeholders or the entire Scrum team, as well as promoting innovation in the development process by employees. A final point is the fact that an agile approach helps to develop realistically. With classic process models, development is a kind of black box for outsiders, as little feedback is given to the developers during development. Agile procedures provide transparency in the development process due to continuous feedback and a continuous target/actual comparison.

Promote collaboration

Agile methods promote collaboration between different roles and stakeholders within the software development process. Agile software development methods encourage an active collaboration between different stakeholders. This can lead to an increase in quality both in the process and the software product itself.

According to the interviews, the developers prefer to work in an agile environment rather than in a classical process model: "*[...] typically, it does not apply to all developers, but the developers prefer to work in this agile model rather than in a classic model. So, for me, the classic model in software development has simply outlived itself in any industry*." *(DATA).* Developing agile entails, a greater appreciation for the developers is shown because they work on their own responsibility and the team is organized independently: "*Many developers say they want to develop agile. That is the motivation. Then we can offer you whatever you need." (INFUSE)*. Active involvement is desired and enabled. This can be transferred from the team level to each team member.

Furthermore, external stakeholders can be involved in the process in Scrum: "*And then we* continued *to pursue this Scrum idea. Even if they're not on the team, just get them closer. So, they will be there on every Scrum-Day. And at the review meeting and see what's new." (INFUSE). "So, if you don't want to join the team and the same office, just get as close as possible and do as much as possible in small steps in the sprint." (XRAY).* Different roles and stakeholders are working closer together. An agile approach also increases employee satisfaction: "*Because the satisfaction is better after all, yes. Employee satisfaction is also important*." *(XRAY).* The team becomes more dynamic, more flexible and more efficient. The human component is a (success) factor and reason to consider agile developments in regulated environments.

To summarize, an agile approach supports working together with different roles and responsibilities. It is, therefore, suitable in a regulated environment as many stakeholders need to be involved and coordinated.

5 Discussion and Conclusion

The goal of this research was to shed light on the benefits and barriers of agile methods used for software development in regulated environments. The aim was not to find new agile practices, but to find the right fit for a regulated compliance context.

Our results confirm that software development in regulated environments emphasize quality assurance, ex-ante planning, additional roles, and documentation. Furthermore, our results show that organizations use agile methods in regulated environments to master complexity, reduce efforts, improve usability, and promote collaboration. Hereafter, we discuss our contributions, limitations and provide avenues for future research.

Our research contributes to the ongoing question why organizations use agile software development methods, given the strict regulations in compliance savvy environments [3]. Our results suggest that agile procedures, such as a project vision, do not only help mastering complexity of both software and hardware solutions that involve APIs and different stakeholders, but also support project planning by separating long-term and short-term planning in different mechanisms. Thereby, our results suggest that a good tailoring of the agile method leaves developers with a suitable methodology to work user-oriented and efficient, meet regulatory requirements [38] [39] as well as perform innovation projects [40].

Our results provide additional detail on how agile methods can be tailored to regulated environments [41]. We identified two different strategies, which we refer to as (1) embedding agile in existing V-model based development and (2) using agile as toolbox of procedures. When embedding agile in an existing plan-based environment, the agile part focuses on the development aspects and usually uses plan-driven approaches for testing and verification. When using agile as a toolbox to improve software development, we found that the teams developed what we refer to as mini-Vs within each sprint: A hybrid approach of combining agile and plan-based elements in each iteration.

Overall, we contribute to the literature on usability of agile methods by enhancing literature on scaling agile methods [42]. Our results provide evidence that agile methods can be successfully applied in regulated environments. We analysed three cases that tailored agile methods to their context by using agile procedures to organize daily work to foster collaboration and focus on user-centricity procedures to ensure usability and cope with complexity. The teams combined this with plan-driven elements for planning and interfaces, especially when complex hardware was involved.

Our study is subject to limitations. First, we selected teams that reported to actively use agile methods and, thus, we might have missed configurations where plan-driven approaches were enriched with agile elements. However, we examined three different cases and asked interviewees about the process, tailoring, and their experiences to ensure a broad perspective. Second, we focused on cases regulated by medical device compliance requirements. We argue that this is a typical regulatory environment but call for future research in different industries such as pharma, banking, and aerospace. Third, the focus of this paper is mostly on the benefits of agile software development in regulated environments. Drawbacks and problems were only touched briefly. We recommend for future research to also consider these aspects. Lastly, it is inherent to exploratory qualitative work that generalizing the results is challenging. For example, we derived our results from agile teams that used Scrum as agile method. Thus, they cannot be taken for granted for methods such as XP or lean. Further research should extend our theoretical sampling and examine other organization size and company context, including distributed teams and other agile methods. Research and practice would benefit from design science research that develops new artifacts, e.g., tools for agile teams in regulated environments or apply method engineering to develop a new development method for applying agile in such environments.

Overall, this research empirically illustrates the tailoring of agile methods for regulated environments. Our findings illustrate that regulated environments emphasize quality assurance, ex-ante planning, additional roles, and documentation. We show that organizations make use of agile methods to master complexity, reduce efforts, improve usability, and promote collaboration.

References

1. Kassab, M., Lee, J., Mazzara, M., Succi, G., Tumyrkin, R.: Software Quality - Traditional vs. Agile: an Empirical Investigation (2016)
2. Hirschmann, C., Dietrich, I., Rauh, D., Wittmann, M.: Scrum im Fluss: Eine Kombination von Scrum und Kanban im regulativen Umfeld. In: OBJEKTspektrum, vol. 06/2014vol. , pp. 49–56 (2014)
3. Fitzgerald, B., Stol, K.-J., O'Sullivan, R., O'Brien, D.: Scaling agile methods to regulated environments. An industry case study. ICSE '13 Proceedings of the 2013 International Conference on Software Engineering, 863–872 (2013)
4. Ambler, S.: Agility at Scale Survey 2016, http://www.ambysoft.com/surveys/agileAtScale2016.html
5. Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates. MDR (2017)
6. Rottier, P.A., Rodrigues, V.: Agile Development in a Medical Device Company. In: Melnik, G. (ed.) Conference Agile, 2008. AGILE '08; August 4 - 8, 2008, Toronto, Ontario, Canada, pp. 218–223. IEEE Computer Soc (2008)
7. Klotz, M.: IT-Compliance. In: Tiemeyer, E. (ed.) Handbuch IT-Management. Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis (2017)
8. Hrgarek, N.: Certification and regulatory challenges in medical device software development. In: 2012 4th International Workshop on Software Engineering in Health Care (SEHC), pp. 40–43. IEEE (2012)
9. McHugh, M., McCaffery, F., Coady, G.: Adopting Agile Practices when Developing Medical Device Software. J Comput Eng Inf Technol 04 (2015)
10. Abrahamsson, P., Conboy, K., Wang, X.: 'Lots done, more to do': the current state of agile systems development research. European Journal of Information Systems 18, 281–284 (2009)
11. Williams, L., Cockburn, A.: Agile software development: it's about feedback and change. Computer 36, 39–43 (2003)
12. Cao, L., Mohan, K., Xu, P., Ramesh, B.: A framework for adapting agile development methodologies. European Journal of Information Systems 18, 332–343 (2009)

13. Holmström, H., Fitzgerald, B., Ågerfalk, P.J., Conchúir, E.Ó.: Agile Practices Reduce Distance in Global Software Development. Information Systems Management 23, 7–18 (2006)
14. Lous, P., Kuhrmann, M., Tell, P.: Is Scrum Fit for Global Software Engineering? In: 2017 IEEE 12th International Conference on Global Software Engineering (ICGSE), pp. 1–10. IEEE (2017)
15. Wortman, K., Duncan, B., Melin, E.: Agile methodology for spacecraft ground software development. A cultural shift. In: 2017 IEEE Aerospace Conference, pp. 1–8. IEEE (2017)
16. Marcal, A.S.C., Soares, F.S.F., Belchior, A.D.: Mapping CMMI Project Management Process Areas to SCRUM Practices. In: Staff, I. (ed.) IEEE/NASA Software Engineering Workshop: 31st: 2007: Columbia, MD, pp. 13–22. IEEE Computer Society Press, [Place of publication not identified] (2007)
17. Cawley, O., Wang, X., Richardson, I.: Lean/Agile Software Development Methodologies in Regulated Environments – State of the Art. In: Abrahamsson, P., Oza, N. (eds.) Lean Enterprise Software and Systems. First International Conference, LESS 2010, Helsinki, Finland, October 17-20, 2010. Proceedings, pp. 31–36. Springer Berlin Heidelberg, Berlin Heidelberg (2010)
18. Laukkarinen, T., Kuusinen, K., Mikkonen, T.: DevOps in Regulated Software Development. Case Medical Devices. In: 2017 IEEE/ACM 39th International Conference on Software Engineering: New Ideas and Emerging Technologies Results Track (ICSE-NIER), pp. 15–18. IEEE (2017)
19. Rasmussen, R., Hughes, T., Jenks, J.R., Skach, J.: Adopting Agile in an FDA Regulated Environment. In: Dubinsky, Y. (ed.) Agile '09, Agile Conference, 2009. 24 - 28 Aug. 2009; Chicago, Illinois, pp. 151–155. IEEE (2009)
20. Jonsson, H., Larsson, S., Punnekkat, S.: Agile Practices in Regulated Railway Software Development. In: 2012 IEEE 23rd International Symposium on Software Reliability Engineering workshops (ISSREW). 27 - 30 Nov. 2012, Dallas, Texas, USA; proceedings, pp. 355–360. IEEE, Piscataway, NJ (2012)
21. Cawley, O., Wang, X., Richardson, I.: Regulated Software Development – An Onerous Transformation. In: Weber, J., Perseil, I. (eds.) Foundations of Health Information Engineering and Systems. Second International Symposium, FHIES 2012, Paris, France, August 27-28, 2012. Revised Selected Papers, 7789, pp. 72–86. Springer Berlin Heidelberg, Berlin/Heidelberg (2013)
22. Shenvi, A.A.: Navigating the maze. In: Janakiram, D., Sen, K., Kulkarni, V. (eds.) Proceedings of the 7th India Software Engineering Conference on - ISEC '14, pp. 1–6. ACM Press, New York, New York, USA (2014)
23. Richtlinie 93/42/EWG vom 14. Juni 1993 über Medizinprodukte. MDD (1993)
24. Medizinproduktegesetz. MPG (1994)
25. IEC 62304: Medizingeräte-Software - Software-Lebenszyklus-Prozesse (2016)
26. ISO 13485:2016-08: Medizinprodukte - Qualitätsmanagementsysteme - Anforderungen für regulatorische Zwecke (ISO 13485:2016)
27. Electronic Code of Federal Regulations. FDA CFR (2018)
28. ISO 14971:2013-04: Medizinprodukte - Anwendung des Risikomanagements auf Medizinprodukte
29. McHugh, M., McCaffery, F., Casey, V.: Barriers to Adopting Agile Practices When Developing Medical Device Software. In: Mas, A., Mesquida, A., Rout, T., O'Connor, R.V., Dorling, A. (eds.) Software Process Improvement and Capability Determination. 12th International Conference, SPICE 2012, Palma, Spain, May 29-31, 2012. Proceedings, 290, pp. 141–147. Berlin, Heidelberg (2012)

30. Przybilla, L., Wiesche, M., Krcmar, H.: The Influence of Agile Practices on Performance in Software Engineering Teams. In: Kishore, R., Beimborn, D., Bandi, R.K., Aubert, B., Compeau, D., Tarafdar, M. (eds.) Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research - SIGMIS-CPR'18, pp. 33–40. ACM Press, New York, New York, USA (2018)
31. Pflügler, C., Wiesche, M., Krcmar, H.: Subgroups in Agile and Traditional IT Project Teams. In: Bui, T. (ed.). Hawaii Int. Conf. on System Sciences (2018)
32. Wiesche, M., Krcmar, H.: The relationship of personality models and development tasks in software engineering. In: Joseph, D., Shayo, C., Guzman, I.R., Kishore, R. (eds.) Proceedings of the 52nd ACM conference on Computers and people research, pp. 149–161. ACM, New York, NY (2014)
33. Rindell, K., Hyrynsalmi, S., Leppanen, V.: Case Study of Security Development in an Agile Environment. Building Identity Management for a Government Agency. In: 2016 11th International Conference on Availability, Reliability and Security. ARES 2016: Salzburg, Austria, 31 August-2 September 2016: proceedings, pp. 556–563. IEEE, Piscataway, NJ, Piscataway, NJ (2016)
34. Yin, R.K.: Case study research. Design and methods. SAGE, Los Angeles, London, New Delhi, Singapore, Washington, DC (2014)
35. Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. VS Verlag, Wiesbaden (2010)
36. Urquhart, C.: Grounded Theory for Qualitative Research. A Practical Guide. Sage Publications (2013)
37. Glaser, B.G., Strauss, A.L.: The discovery of grounded theory. Strategies for qualitative research. Aldine, New Brunswick (2009)
38. Fitzgerald, B., Russo, N.L., O'Kane, T.: Software development method tailoring at Motorola. Commun. ACM 46, 64–70 (2003)
39. Recker, J., Holten, R., Hummel, M., Rosenkranz, C.: How Agile Practices Impact Customer Responsiveness and Development Success: A Field Study. Project Management Journal 48, 99–121 (2017)
40. Przybilla, L., Schreieck, M., Klinker, K., Wiesche, M., Krcmar, H.: Combining Design Thinking and Agile Development to Master Highly Innovative IT Projects. Projektmanagement und Vorgehensmodelle 2018. In: Gesellschaft für Informatik e. V. (GI) (ed.) Lecture Notes in Informatics (LNI), 2018. Bonn (2018)
41. Hajou, A., Batenburg, R.S., Jansen, S.: Method æ, the Agile Software Development Method Tailored for the Pharmaceutical Industry. LNSE 3, 251–262 (2015)
42. Goh, J., Pan, S., Zuo, M.: Developing the Agile IS Development Practices in Large-Scale IT Projects: The Trust-Mediated Organizational Controls and IT Project Team Capabilities Perspectives. JAIS 14, 722–756 (2013)

Are Employees Following the Rules? On the Effectiveness of IT Consumerization Policies

Michael Klesel[1,2], Sebastian Weber[1], Finja Walsdorff[1], and Bjoern Niehaves[1]

[1] University of Siegen, Siegen, Germany
{michael.klesel,sebastian.weber,finja.walsdorff,bjoern.niehaves}
@uni-siegen.de
[2] University of Twente, Enschede, The Netherlands

Abstract. In most organizations, employees commonly use mobile technologies including smartphones and tablets to complete their tasks. Therefore, many organizations have started to implement policies that govern the use of mobile devices such as Bring-Your-Own-Device (BYOD) policies, that allow employees to use private devices for work-related purposes, or Company Owned Privately Enabled (COPE) policies, which allow the use of organizational technologies for private purposes. Despite its relevance, there is only little empirical research that provides evidence on the effectiveness of specific policies, i.e., policies in favor of BYOD/COPE, policies that prohibit it, and no implemented policies. Based on survey data ($N = 381$), we provide initial insights in terms of the effectiveness of these policies. Our results indicate that policies indeed influence the degree of technology use. Policies in favor of BYOD/COPE are particularly effective. We conclude this paper by discussing our findings and derive several implications for theory and practice.

Keywords: IT Consumerization, BYOD, COPE, policies

1 Introduction

In modern organizations, it has become common practice for employees to use their personal technological devices or applications for job-related tasks. This use of consumer IT in the workplace, also known as "IT consumerization", represents a significant change in contemporary work life and has different benefits such as increased innovation, enhanced productivity, and a higher level of employee satisfaction [1]. A maturing body of knowledge has contributed to a better understanding of this development by investigating specific antecedents of IT consumerization [2] and its effects [3].

Although literature on IT consumerization has matured, it hardly investigates organizational aspects. This gap becomes most evident with regard to the effects of policy implementation and its impact on use behavior (see for instance [4]). From a practical perspective, this shortcoming is quite significant because policies are powerful instruments that allow organizations to influence their employees' use behavior (e.g., [5]). Therefore, it is important to provide empirical evidence on the usefulness and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

effectiveness of policies to derive implications and consult with organizations. Since there are two dominant options, namely Bring Your Own Device (BYOD) strategies [6] and Company Owned Privately Enabled (COPE) strategies [7], which are both equally important for organizations, a detailed analysis of their effectiveness would be most beneficial. Against this background, we seek to address this aspect with the following research questions (RQ):

RQ 1: How effective are policies in enhancing/reducing device use behavior?
RQ 2: Is there a difference between the effectiveness of BYOD and COPE policies?

In order to address our research questions, this paper is structured as follows: first, we review existing literature on IT consumerization to identify existing knowledge on the effects of policies in this field (section 2). Based on our review, we propose a research model that addresses this issue by hypothesizing differences in terms of policies and their impact on use behavior (section 3). In Section 4, we describe our methodological approach to address our objective. Thereafter, we present our analysis and the results (section 5). We discuss our findings in Section 6 and conclude by reflecting on the limitations of our study and by providing impulses for future research (section 7).

2 Related Work on IT Consumerization

Today, it is common practice to use privately-owned technologies for work-related tasks or to use company-provided IT for private purposes. The blend of personal technological devices or applications and business IT is described as "IT consumerization" or "consumerization of IT" [8]. This development comes with distinct strategies that allow organizations to monitor this phenomenon: BYOD and COPE. While BYOD refers to employees' work-related use of private hardware devices (such as personal laptops, smartphones, or tablets), COPE refers to the personal use of corporate IT. A typical example of COPE is the dual use of company-provided technologies such as smartphones for professional and private communication. In contrast, using a privately-owned smartphone for both purposes is considered BYOD.

Literature on IT consumerization acknowledges the multiple perspectives on this phenomenon. For example, Harris et al. [1] and Köffer et al. [6] refer to three perspectives, namely the individual, the organizational and the market perspective. The individual perspective refers to how an individual handles personal IT that is brought to work and is used for work-related purposes [1]. The individual perspective on IT consumerization thus focuses on the ownership of an IT tool [6]. The organizational perspective on the other hand deals with governing the use of such private IT in official business settings. From this point of view, IT consumerization can be seen as both a threat and an opportunity [1]. Finally, the market perspective on IT consumerization focuses on the origin or target market of consumer IT [1]. This third perspective highlights that consumer market technologies gradually reach enterprises, thereby having an impact on the IT department and preventing the distinction between consumer and enterprise IT [6]. It is worth noting that those perspectives overlap and influence each other.

In order to identify literature that addresses our research questions, we conducted a structured literature review [9, 10]. We started with the Web of Science by searching for "IT consumerization" within the senior scholar basket of eight [11]. Moreover, we manually scanned the forthcoming section of each journal. This means that most current publications [12, 13] are also considered here. Since the initial search did only reveal two studies, we further consulted the AIS library, which yielded in 24 papers. Note that the search was conducted by means of a keyword search in the abstract without limitations regarding the publication year. Each of the 26 papers was read by at least one of the authors and classified into one of the three perspectives. We specifically took all research questions, the research design, and the data sample into consideration and analyzed the unit of analysis. Since the three perspectives are not distinct (i.e., they overlap), we used the dominant perspective as a criterion to which we assigned each paper. An overview is given in Table 1.

For instance, Junglas et al. [13] seek to "examine the effect of IT consumerization on innovative work behaviours" (p. 2). Since the main focus lies on the individual, we included this study in the individual perspective. Similarly, we included research that investigates the change of governance structures [12] in the organizational section. Studies that examine the market perspective such as Niehaves et al. [8] ("[w]hat areas of information systems are specifically affected by consumerization", p. 2) were categorized accordingly.

Table 1. Related work on IT Consumerization

Perspective	Focus	Typical research question	References
Market	"Origin or intended target market of the IT tool" ([6], p. 366)	"What areas of information systems are specifically affected by consumerization?" ([8], p. 2)	[8]
Individual	"Ownership of the IT tool" ([6], p. 366)	"Why do some employees choose novel and innovative consumer IT on their own while others continue to work with the existing enterprise IT?" ([14], p. 1)	[2, 3, 6, 13–23]
Organizational	"Permission to use private IT tools for work" ([6], p. 366)	"What conflicts does IT consumerization create for IT departments?" ([24], p. 4)	[1, 4, 12, 24–31]

Our review highlights that the primary focus of research on IT consumerization lies on the individual perspective, whereas the market perspective only seems to play a minor role. This indicates that the origin of an IT tool, which is the focal point of the market perspective, is not of crucial importance in current literature anymore. This trend seems rather natural due to the pervasive nature of consumer technologies today. In turn, this makes the distinction between consumer market and business market less important. Our review also shows that, while the organizational perspective on IT consumerization has expanded in general, there is not yet a lot of literature that addresses the subject of policies. If at all, policies in relation to IT consumerization are of secondary importance in the reviewed literature. Notable examples include a study conducted by Lüker et al. [4] who make an important contribution to the field by investigating the importance of IT consumerization policies. Similarly, Mokosch et al. [28] study how organizational structures affect individual behavior. This lack of research is very significant because policies are a fundamental aspect of workplace design and, therefore, influence individual behavior. Without empirical evidence that provides further insights into this relationship, it is challenging to justify the implementation of such policies. Against this background, we seek to expand existing knowledge on the role of policies by investigating the impact of different types of policies on individual use behavior based on a large empirical sample. Our underlying research model is proposed in the following section.

3 Research Model

It is commonly accepted that external factors such as organizational factors including policies have a major impact on technology use. This is a fundamental assumption in the technology acceptance literature (e.g., [32]) and has also been emphasized in literature on mobile devices [28, 33]. The link between policies and individual behavior has also received support in various fields. For instance, Richman et al. [34] show that there is a significant relationship between supportive work-life policies and an employee's engagement. Similarly, Moskowitz et al. [35] highlight that workplace smoking policies result in employees smoking less.

Surprisingly, in related literature on IT Consumerization, the link between policies and use behavior is somewhat ambiguous. On the one hand, there is an increasing amount of literature that shows that employees use their devices despite it being prohibited or at least not approved. This trend is commonly referred to as shadow IT (e.g., [36]). In fact, existing research suggests that individuals who seek to meet their job performance expectations willingly neglect IS policies [37]. Based on this stream of research, we can assume that policies might not yield the intended degree of effectiveness. On the other hand, current literature provides evidence that the type of policy indeed has a significant effect on use behavior. For example, Junglas et al. [13] show a significant relationship between permission to use and intended use behavior.

In order to investigate the impact of different policies in detail, we propose a research model that explicitly differentiates between three distinct types of policies: policies that allow BYOD/COPE, policies that prohibit BYOD/COPE, or a lack of policy (i.e., no policies are implemented).

Policies that allow BYOD/COPE. We generally assume that policies allowing BYOD/COPE have a positive impact on use behavior. This is in line with existing literature on IT Consumerization (e.g., [13, 28]). We specifically assume that a policy that is in favor of BYOD or COPE has a positive impact on the corresponding degree of technology use. In fact, current literature suggests that permission to use private technology at the workplace has a significant impact on the individual's decision to actually use it [13]. We assume that this relationship is stronger when organizations have a policy in place that is in favor of BYOD/COPE rather than lacking such a guideline. Therefore, we propose the following hypotheses:

Hypothesis 1a:	A policy that allows BYOD leads to a higher degree of (private) device use (for work-related purposes) than a lack of policy.
Hypothesis 1b:	A policy that allows COPE leads to a higher degree of (company-owned) device use (for private purposes) than a lack of policy.

In line with hypothesis 1a and 1b, we assume that there also is a significant difference between organizations that allow BYOD/COPE and organizations that prohibit it. Therefore, we hypothesize the following:

Hypothesis 2a:	A policy that allows BYOD leads to a higher degree of (private) device use (for work-related purposes) than a policy that prohibits BYOD.
Hypothesis 2b:	A policy that allows COPE leads to a higher degree of (company-owned) device use (for private purposes) than a policy that prohibits COPE.

Policies that prohibit BYOD/COPE. Based on the IT consumerization literature that investigates the prevention of security threats and non-compliant behavior (e.g., [4, 38]), prohibition policies are of major interest. Those studies suggest that awareness of prohibition policies as well as the possible cost of noncompliance effectively decrease the use of technology. This is also in line with more general literature on security compliance relating to deterrence theory. This theory suggests that the higher the cost of noncompliance (e.g., sanctions, privacy concerns), the more likely it is that employees comply with given policies (e.g., [39, 40]). Contrary to this assumption, literature on shadow IT proposes that employees tend to be pragmatic and care more about their job performance than about complying with IS policies [37]. Those types of behavior can be explained by the neutralization theory, which suggests that people use neutralization techniques to legitimize their misbehavior [41]. For instance, Silic et al. [42] show that the neutralization technique "metaphor of ledger" has a strong and positive effect on policy violation in the context of shadow IT. This technique is based on the idea that we compensate bad behavior (e.g. violating the policy) with good behavior (e.g., overtime) [42]. Considering both streams of research, we argue that the existence of a prohibition policy has a stronger effect than a lack of policy. Hence, we propose the following:

Hypothesis 3a:	A policy that prohibits BYOD leads to a lower degree of (private) device use (for work-related purposes) than a lack of policy.
Hypothesis 3b:	A policy that prohibits COPE leads to a lower degree of (company-owned) device use (for private purposes) than a lack of policy.

Combining the arguments mentioned above, we propose a model that compares different relationships between policies that govern different types of use behavior and the actual use (cf. Figure 1).

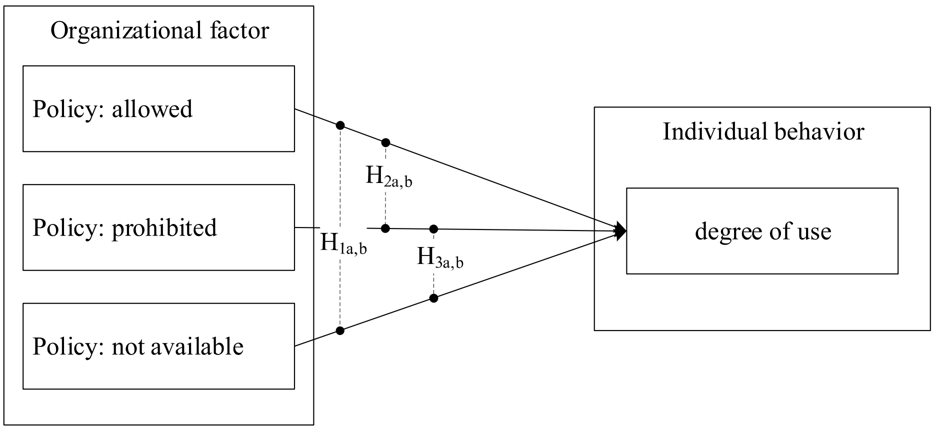


Figure 1. Research model

4 Methodology

This study is part of a larger project on IT consumerization and its impact on organizations. The focus of this study is related to the role of organizational policies and how they affect individual behavior. We collected data from 400 employees using computer-assisted telephone interviews (CATI). The participants were recruited from different local administrations in Germany that ranged from less than 50 employees up to 10,000. In order to create a representative sample, we selected 400 administrations out of the overall administration population. Responses with missing values relating to policies or use behavior were excluded, which yields 381 usable observations for this study. Table 2 provides a summary of the demographics. To measure the perceived skill in terms of technology use, the participants rated their IT skills ("How would you rate your IT skills…") on an ordinal scale ranging from "beginner" to "competent user" to "expert".

Table 2. Demographics, $N = 381$

Dimension	Classification	Percentage
Age	20-39	21.3%
	40-60	61.2%
	Older than 60	11.5%
	n/a	6.0%
Gender	male	67.4%
	female	32.6%
Skills	beginner	20.5%
	competent user	60.1%
	expert	19.1%
	n/a	0.3%

Whenever possible, we used established measurements. For BYOD and COPE use, we applied the device-related dimension of an established scale [2][1]. In terms of policies, we adopted an existing categorical [2] including "not allowed", "allowed", "missing". An overview is given in Table 3:

Table 3. Measurement Items

	Variable	Adapted Item	Scale	Source
BYOD	Use	I use private devices (e.g., laptop, smartphone) to complete work tasks.	7-point Likert scale	(adopted from [2])
	Policy	In my organization, the use of private devices to perform work tasks is …	"not allowed", "allowed", "missing"	(adopted from [13])
COPE	Use	I use the devices provided by the organization (e.g., laptop, smartphone) to complete private tasks.	7-point Likert scale	(adopted from [2])
	Policy	In my organization, the use of devices provided by the organization to perform private tasks is …	"not allowed", "allowed", "missing"	(adopted from [13])

[1] We also computed the analysis with the complete scale, including the use of internet accounts and the use of software [2] with sum scores. Since the results did not yield significantly different results, we only use one item for the subsequent analysis. Therefore, it is more in line with the objective of this study.

5 Data Analysis and Results

In order to investigate the hypotheses, the data analysis for each policy (i.e., BYOD and COPE) was conducted in three consecutive steps.

First, we applied a multiple regression in order to identify possible confounding factors. Therefore, we included gender, age, and skills as possible factors (i.e., predictors) of use in our regression models because these are commonly included in use-related research (e.g., [43]). The results of the BYOD regression model are significant in that they predict use and show that the three predictors account for 2.5% of the variance, $R^2 = .025$, F (3, 375) = 3.216, $p < .023$. We found no significant predictive power for the coefficients of age, $\beta = .063$, $p = .223$, and skills, $\beta = .057$, $p = .261$. However, the data shows that gender is significant, $\beta = -.124$, $p < .016$. The regression for COPE yielded similar results. The regression model shows that the three predictors account for 4.3% of the variance, $R^2 = .043$, F (3, 377) = 5.666, $p < .001$. Age, $\beta = .022$, $p = .663$, and skills, $\beta = -.021$, $p = .679$, do not significantly predict use while gender contributes to the model significantly, $\beta = -.020$, $p < .001$. Thus, the results indicate that gender has a confounding effect. Consequently, we include gender in our consecutive analysis.

Second, to analyze the effects of policies on the use of consumer IT, we applied an analysis of covariance (ANCOVA) including gender as a covariate. Table 4 shows the results that indicate a significant effect of BYOD policy on BYOD use after controlling for gender, F (2, 377) = 13.93, $p < .001$.

Table 4. ANCOVA results and Descriptive Statistics for BYOD Use by Policy and Gender

Type of Policy	BYOD Use			
	M	Adj. *M*	*SD*	*n*
Use prohibited	1.91	1.89	1.61	138
No Regulation	2.26	2.28	1.76	172
Use allowed	3.21	3.21	1.90	71
Source	*SS*	*df*	*MS*	*F*
Gender	24.44	1	24.44	8.26*
Policy	82.42	2	41.21	13.93**
Error	1115.58	377	2.96	

Note. *$p < .01$, ** $p < .001$

Similarly, Table 5 shows the significant effect of COPE policy on COPE use after controlling for gender, F (2, 377) = 17.43, $p < .001$.

Third, due to the statistically significant results, we carried out post hoc comparison analyses using Tukey's honestly significant difference (HSD) test to further examine the differences between the policies. For BYOD, the post hoc Tukey tests show that the "use allowed" policy ($M = 3.21$), no regulation ($M = 2.26$) and the "use prohibited" policy ($M = 1.91$) differ significantly at p < .01; the "use prohibited" policy and "no regulation" were not significantly different (see Figure 2). For COPE, the post hoc Tukey tests yield similar results. The "use allowed" policy ($M = 2.49$) differs significantly at p < .01 compared to "no regulation" ($M = 1.63$) and the "use prohibited" policy ($M = 1.40$); the "use prohibited" policy and "no regulation" were not significantly different (see Figure 2).

Table 5. ANCOVA Results and Descriptive Statistics for COPE Use by Policy and Gender

Type of Policy	COPE Use			
	M	Adj. *M*	*SD*	*n*
Use prohibited	1.40	1.41	1.11	195
No Regulation	1.63	1.64	1.21	116
Use allowed	2.49	2.43	1.64	70
Source	*SS*	*df*	*MS*	*F*
Gender	19.43	1	19.43	12.81**
Policy	52.88	2	26.44	17.43**
Error	571.91	377	1.52	

Note. $*p < .01$, $**\, p < .001$

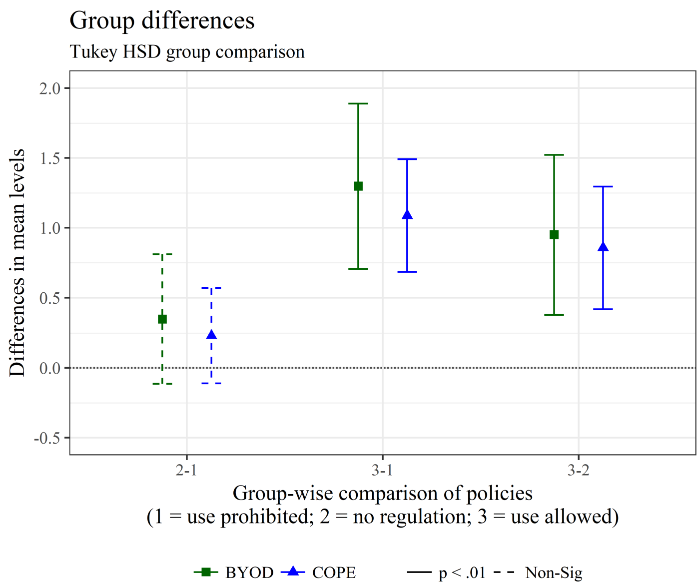


Figure 2. Group-wise differences

6 Discussion

Overall, our results indicate that BYOD and COPE policies can be compared in terms of their impact on individual use behavior. Therefore, our results hold for both types. We do recognize small differences between BYOD/COPE regarding their mean levels.

Specifically speaking, the differences between COPE policy groups are smaller than the differences between BYOD policy groups (see Figure 2). This indicates that the effect of BYOD policies is slightly stronger in terms of absolute values. We also conclude that the use of private devices for work related tasks (BYOD) is more common than using company owned devices for private purposes (COPE). In comparison, each BYOD policy type (i.e., "use prohibited", "no regulation", "use allowed") has a higher mean use than COPE policies (see Table 4 and Table 5).

The results also suggest that the hypotheses that a policy that allows BYOD/COPE leads to a higher degree of device use than a lack of policy (H1a, H1b) or a policy that prohibits BYOD/COPE (H2a, H2b) can both be supported. Thus, we can conclude that the implementation of policies in favor of using technology is an important tool for organizations to promote use. However, H3a and H3b, which are proposing that a policy that prohibits BYOD/COPE yields a lower degree of use compared to a lack of policy have to be rejected because no significant differences were detected. This does not necessarily imply that policies that prohibit use are not effective. Based on our data, this is rather due to a generally low level of use (BYOD – *M*: 1.91; COPE – *M*: 1.40). Consequently, we argue that no regulation reduces device use. As hypothesized in H1a and H1b, perceived risks may have an influence on whether a device is used or not in the case of no regulation. Since our findings were focused on the overall effect of policies and not on the individual antecedents of device use, we did not measure those factors. But we believe that those perceived risks are a valid explanation for these results.

Based on our insights, we derive several implications for theorizing. Most importantly, our results suggest that there are situations where there are no differences between an implemented policy and a lack of policy. Against this background, existing IT consumerization studies on policies could be examined in more detail. For instance, Lüker et al. analyze a compliant behavior related to different specified prohibition policies (loose vs. strict) [4]. Since they do not explicitly distinguish between a lack of policy and prohibition policies, their research can benefit from our results. Similarly, Junglas et al. show that policies have a significant impact on IT consumerization behavior [13]. Again, their study could be further expanded by investigating different types of policies.

Since the primary objective of this study is to investigate the effectiveness of different policies, this research has important implications for organizations. Most importantly, this study provides empirical evidence that the implementation of policies is effective. Therefore, organizations that are interested in increasing (or decreasing) their staff's use behavior (such as BYOD) can implement policies that follow their intended strategy. This is also in line with previous literature that investigates the relationship between organizational structures and individual behavior [28].

Based on our results, there is no significant difference between policies that prohibit use and a lack of policy. Hence, if organizations want to benefit from advantages related to consumer IT (e.g., innovation behavior [6] or performance [3]) they should implement policies that explicitly allow the use of private devices.

7 Conclusion, Limitations and Outlook

Based on a comprehensive review of the IT Consumerization literature, this paper addresses a gap by examining the effectiveness of distinct BYOD and COPE policies. The results show that policies generally are an important tool to influence the degree of use regarding BYOD and COPE. Above all, they highlight that if an organization is interested in benefitting from consumer IT (e.g., innovation behavior [6] or performance [3]), they should implement policies in favor of BYOD and COPE.

Not unlike any other study, this piece of research has its limitations, which, in turn, opens the door for future research. First, our results suggest that devices are used rather sparingly (see mean values). A possible explanation for this is our sample. In fact, there is a high proportion of older people (see Table 2). Since literature on digital divide suggests that older people play a pivotal role (e.g., [44]), future research should shed further light on the role of age. Second, this research builds on survey data. Based on our findings, further research could extend these efforts in terms of field experiments in order to go beyond self-reported indications. Third, our literature analysis is based on previous work on "IT Consumerization". However, there is more literature available that can be included by extending the keywords. For example, "Individual Information Systems" can be included to get a more comprehensive overview. Fourth, there might be situations where organizations force their employees to use private IT (e.g., gig economy). Hence, future research should also investigate how enforcement affects the degree of individual use behavior. Finally, the role of policies may differ across specific groups. Therefore, we suggest to further investigate this topic by acknowledging group differences such as industry versus government or large organizations versus small and medium sized enterprises.

8 Acknowledgements

We would like to acknowledge that this research was supported by the National E-Government Competence Center (NEGZ) and the German Federal Ministry of Education and Research (BMBF, promotional reference 02L14A011).

References

1. Harris, J., Ives, B., Junglas, I.: IT Consumerization: When Gadgets Turn into Enterprise IT Tools. MIS Q. Exec. 11, 99–112 (2012).
2. Ortbach, K., Köffer, S., Bode, M., Niehaves, B.: Individualization of Information Systems - Analyzing Antecedents of IT Consumerization Behavior. Presented at the Proceedings of the 34th International Conference on Information Systems (ICIS) , Milan, Italy (2013).
3. Köffer, S., Ortbach, K., Niehaves, B.: Exploring the Relationship between IT Consumerization and Job Performance: A Theoretical Framework for Future Research. Commun. Assoc. Inf. Syst. 35, 261–283 (2014).

4. Lüker, N., Winkler, T., Kude, T.: IT Consumerization and Compliant Use. Presented at the Proceedings of the 20th Pacific Asia Conference on Information Systems (PACIS) , Chiayi, Taiwan (2016).
5. Vardi, Y., Weitz, E.: Misbehavior in Organizations: Theory, Research, and Management. L. Erlbaum, Mahwah, N.J (2004).
6. Köffer, S., Ortbach, K., Junglas, I., Niehaves, B., Harris, J.: Innovation through BYOD? The Influence of IT Consumerization on Individual IT Innovation Behavior. Bus. Inf. Syst. Eng. 57, 363–375 (2015).
7. Rathnasekara, C., Athukorala, T., Dikwellage, L., Wickramasuriya, U., Senarathne, A., Elvitigala, S.: Securing Corporate Data in Mobile Devices in a COPE Environment. Presented at the 6th National Conference on Technology and Management (NCTM) , Malabe, Sri Lanka January (2017).
8. Niehaves, B., Köffer, S., Ortbach, K.: IT Consumerization – A Theory and Practice Review. Presented at the Proceedings of the 18th Americas Conference on Information Systems (AMCIS) , Seattle, Washington (2012).
9. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, xiii–xxiii (2002).
10. Boell, S.K., Cecez-Kecmanovic, D.: On Being "Systematic" in Literature Reviews in IS. J. Inf. Technol. 30, 161–173 (2015).
11. AIS: Senior Scholars' Basket of Journals, https://aisnet.org/?SeniorScholarBasket. (Accessed: 18.12.2018)
12. Gregory, R.W., Kaganer, E., Henfridsson, O., Ruch, T.J.: IT Consumerization and the Transformation of IT Governance. MIS Q. 42, 1225–1253 (2018).
13. Junglas, I., Goel, L., Ives, B., Harris, J.: Innovation at Work: The Relative Advantage of Using Consumer IT in the Workplace. Inf. Syst. J. in press, 1–23 (2018).
14. Ortbach, K., Bode, M., Niehaves, B.: What Influences Technological Individualization? – An Analysis of Antecedents to IT Consumerization Behavior. Presented at the Proceedings of the 19th Americas Conference on Information Systems (AMCIS) , Chicago, Illinois (2013).
15. Buettner, R., Buettner, R.: Towards a New Personal Information Technology Acceptance Model: Conceptualization and Empirical Evidence from a Bring Your Own Device Dataset. Presented at the Proceedings of the 21st Americas Conference on Information Systems (AMCIS) , San Juan, Puerto Rico (2015).
16. Junglas, I., Goel, L., Ives, B., Harris, J.: Consumer IT at Work: Development and Test of an IT Empowerment Model. Presented at the Proceedings of the 35th International Conference on Information Systems (ICIS) , Auckland, New Zealand (2014).
17. Mueller, M., Klesel, M., Heger, O., Niehaves, B.: Empirical Insights on Individual Innovation Behaviour: A Qualitative Study on IT-Consumerization. Presented at the Proceedings of the 20th Pacific Asia Conference on Information Systems (PACIS) , Chiayi, Taiwan (2016).
18. Niehaves, B., Köffer, S., Ortbach, K.: The Effect of Private IT Use on Work Performance - Towards an IT Consumerization Theory. In: Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013). pp. 39–54. , Leipzig, Germany (2013).
19. Niehaves, B., Ortbach, K., Köffer, S., Reimler, S.: Boon and Bane of IT Consumerization: The Burnout-Engagement-Continuum. Presented at the Proceedings of the 19th Americas Conference on Information Systems (AMCIS) , Chicago, Illinois (2013).
20. Ostermann, U.: What Drives Employees to Become Active? Presented at the Proceedings of the 21st DIGIT Workshop , Dublin, Ireland (2016).

21. Pirani, N., Meister, D.: IT Consumerization: A Model of Private IT Use in Organizations. Presented at the Proceedings of the 19th DIGIT Workshop , Auckland, New Zealand (2014).
22. Weeger, A., Gewald, H.: Factors Influencing Future Employees' Decision-Making to Participate in a BYOD Program: Does Risk Matter? Presented at the Proceedings of the 22nd European Conference on Information Systems (ECIS) , Tel Aviv, Israel (2014).
23. Zaza, I., Armstrong, D.: A Look on the Generational Differences in IT Self-Service Engagement. Presented at the Proceedings of the 24th Americas Conference on Information Systems (AMCIS) , New Orleans, Louisiana (2018).
24. Koch, H., Zhang, S., Giddens, L., Milic, N., Yan, K., Curry, P.: Consumerization and IT Department Conflict. Presented at the Proceedings of the 35th International Conference on Information Systems (ICIS) , Auckland, New Zealand (2014).
25. Yan, K., Zhang, S., Milic, N., Koch, H., Curry, P.: IT Consumerization and New IT Practices: Discriminating, Firefighting and Innovating. Presented at the Proceedings of the 22nd Americas Conference on Information Systems (AMCIS) , San Diego, California (2016).
26. Ahuja, S., Gallupe, B.: A Foundation for the Study of Personal Cloud Computing in Organizations. Presented at the Proceedings of the 21st Americas Conference on Information Systems (AMCIS) , San Juan, Puerto Rico (2015).
27. Köffer, S., Fielt, E., Niehaves, B.: IT Consumerization and its Effects on IT Business Value, IT Capabilities, and the IT Function. Presented at the Proceedings of the 19th Pacific Asia Conference on Information Systems (PACIS) , Singapore, Singapore (2015).
28. Klesel, M., Mokosch, G., Niehaves, B.: Putting Flesh on the Duality of Structure: the Case of IT Consumerization. Presented at the Proceedings of the 21st Americas Conference on Information Systems (AMCIS) , San Juan, Puerto Rico (2015).
29. Ortbach, K., Walter, N., Öksüz, A.: Are You Ready to Lose Control? A Theory on the Role of Trust and Risk Perception on Bring-Your-Own-Device Policy and Information System Service Quality. Presented at the Proceedings of the 23rd European Conference on Information Systems (ECIS) , Münster, Germany (2015).
30. Ortbach, K., Köffer, S., Müller, C.P.F., Niehaves, B.: How IT Consumerization Affects the Stress Level at Work: A Public Sector Case Study. Presented at the Proceedings of the 17th Pacific Asia Conference on Information Systems (PACIS) , Jeju Island, Korea (2013).
31. Ortbach, K., Brockmann, T., Stieglitz, S.: Drivers for the Adoption of Mobile Device Management in Organizations. Presented at the Proceedings of the 22nd European Conference on Information Systems (ECIS) , Tel Aviv, Israel (2014).
32. Burton-Jones, A., Hubona, G.S.: The Mediation of External Variables in the Technology Acceptance Model. Inf. Manage. 43, 706–717 (2006).
33. Jarvenpaa, S.L., Lang, K.R.: Managing the Paradoxes of Mobile Technology. Inf. Syst. Manag. 22, 7–23 (2005).
34. Richman, A.L., Civian, J.T., Shannon, L.L., Jeffrey Hill, E., Brennan, R.T.: The Relationship of Perceived Flexibility, Supportive Work–Life Policies, and Use of Formal Flexible Arrangements and Occasional Flexibility to Employee Engagement and Expected Retention. Community Work Fam. 11, 183–197 (2008).
35. Moskowitz, J.M., Lin, Z., Hudes, E.S.: The Impact of Workplace Smoking Ordinances in California on Smoking Cessation. Am. J. Public Health. 90, 757–761 (2000).
36. Haag, S., Eckhardt, A.: Shadow IT. Bus. Inf. Syst. Eng. 59, 469–473 (2017).
37. Guo, K.H., Yuan, Y., Archer, N.P., Connelly, C.E.: Understanding Nonmalicious Security Violations in the Workplace: A Composite Behavior Model. J. Manag. Inf. Syst. 28, 203–236 (2011).

38. Crossler, R.E., Long, J.H., Loraas, T.M., Trinkle, B.S.: Understanding Compliance with Bring Your Own Device Policies Utilizing Protection Motivation Theory: Bridging the Intention-Behavior Gap. J. Inf. Syst. 28, 209–226 (2014).
39. D'Arcy, J., Hovav, A., Galletta, D.: User Awareness of Security Countermeasures and its Impact on Information Systems Misuse: A Deterrence Approach. Inf. Syst. Res. 20, 79–98 (2009).
40. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. MIS Q. 34, 523–548 (2010).
41. Siponen, M., Vance, A.: Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations. MIS Q. 34, 487–502 (2010).
42. Silic, M., Barlow, J.B., Back, A.: A new Perspective on Neutralization and Deterrence: Predicting Shadow IT Usage. Inf. Manage. 54, 1023–1037 (2017).
43. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Q. 27, 425–478 (2003).
44. Niehaves, B., Plattfaut, R.: Internet Adoption by the Elderly: Employing IS Technology Acceptance Theories for Understanding the Age-Related Digital Divide. Eur. J. Inf. Syst. 23, 708–726 (2014).

Agile and Attached: The Impact of Agile Practices on Agile Team Members' Affective Organisational Commitment

Barbara Prommegger[1], Veronika Huck-Fries[1], Manuel Wiesche[1], Helmut Krcmar[1]

[1]Technical University of Munich, Department of Informatics, Munich, Germany
{barbara.prommegger,veronika.fries,wiesche,krcmar}@in.tum.de

Abstract. The current shortage of information systems (IS) specialists is leading to a strongly competitive labour market for the IT workforce. Technology companies need opportunities to prevent high replacement costs and knowledge loss by strengthening the affective organisational commitment (affective OC) of their employees. Using structural equation modelling, we investigate the influence of agile information systems development (ISD) on team members' affective OC. Our results demonstrate that agile project management (APM) positively predicts affective OC directly as well as indirectly via team members' job autonomy (JA) and their supervisors' support (SS). Our study gives empirical evidence on the relationship between agile ISD practices and affective OC and provides implications how to successfully leverage team members' affective OC. For practitioners, our research expounds why and how agile ISD is a suitable instrument to transform leadership culture within the company so as to raise affective OC beyond the IT workforce.

Keywords: Agile information systems development, affective organisational commitment, IT workforce, supervisor support, job autonomy

1 Introduction

Expenditure on IT is currently at its highest level and points to a new cycle of growth in the IT industry [1]. Consequently, the IT sector is now the strongest sector with respect to jobs in Germany, which is why IT professionals are in greater demand than ever before [2]. Simultaneously, the IT industry suffers from high labour fluctuation [3]. A current average turnover rate of 13.2% makes the software industry the unfortunate leader among all industries in this regard [4]. The consequences of turnover, such as high replacement costs and a significant loss of knowledge [3], indicate the need for measures that counteract the resignation of employees.

One main factor influencing turnover is organisational commitment (OC). Committed employees feel personally attached to their company [5], and therefore, they are willing to invest effort on behalf of the organisation [6]. To bind their employees, companies have to be attractive and convey values that employees regard as positive [7]. By introducing and living a culture within the company to which employees can

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

relate, they (the employees) identify with the organisation and satisfaction, engagement and performance increase [6].

Currently, a widely used method of sustainably modifying not only work processes but also values in the organisation is the introduction of agile information systems development (ISD). Using the concept of job redesigning, agile ISD practices influence job characteristics, as well as value constructs, and thereby, they change individuals' outcomes in terms of job satisfaction and motivation [8, 9]. We use this perspective of agility as a form of job redesigning to investigate the connection between agile practices and attachment by examining this question: "*What are the effects of agile project management (APM) practices on the affective OC of agile ISD team members*?"

We claim that agile ISD practices predict affective OC both directly and indirectly. First, agile practices interfere with values embedded within the organisation [10, 11], and consequently, they affect the value congruence and the identification of employees with the organisation. Second, by following the principle of self-organisation, agile ISD shifts responsibility to the team level, and at the same time, it provides a supervisor with the ability to support and protect the team in the face of difficulties [12, 13]. This rise of autonomy and supervisor support determines the commitment of employees to an organisation [14]. Using structural equation modelling, we aim to demonstrate the impact of agility both directly and indirectly on affective OC.

The paper is structured as follows. First, we explain the theoretical background research regarding agile ISD as well as affective OC. Second, we introduce the research model, as well as the hypotheses based on the literature review. Third, we describe the research method, as well as the study approach, and introduce the results of the structural equation model. Last, we discuss the implications and give further recommendations concerning affective OC in the context of agile ISD.

2 Theoretical Background

2.1 Agile Information Systems Development

Agile ISD is widely used in companies of various sizes and those belonging to various industries [15]. Within our paper, we define agile ISD as "the continual readiness of an information systems development method to rapidly or inherently create change, proactively or reactively embrace change, and learn from change while contributing to perceived customer value (economy, quality, and simplicity), through its collective components and relationships with its environment" [16]. The strength of agile ISD practices is decisive here: Agile approaches react to the challenges and needs of a fast-moving environment. By embracing change in projects, agility enables flexible planning, even when the scope of projects remains unstable [17].

At the employee level, agile ISD acts as an instrument for job design [8]. Agility improves team structures [18], reduces negative effects of subgroups [19] and influences job characteristics such as skill variety and task identity [8]. Furthermore, the ability to complete a whole task in the form of "user stories" leads to an increase in motivation, as well as ambition, to improve constantly [9]. Because of these factors, agile ISD influences the working processes of agile team members.

Agile ISD also causes an alteration in individuals' values and principles, changing the software development process fundamentally. In contrast to traditional development methods, agile ISD puts the focus on people and collaboration [11, 20]. Agile ISD trusts in self-organised teams that decide independently the extent and implementation of the requirements [21]. Equally, the supervisor assumes a coordinating and mentoring role [22]. This shift to flat hierarchies reduces cumbersome processes and makes agile teams faster and more efficient, allowing high-speed and top performance within the software development process [12, 13].

In our paper, we define an agile ISD team as a cross-functional team building and updating software, mostly using agile ISD practices and including members both focusing on delivering software as well as managing the team [8].

2.2 Affective Organisational Commitment

OC is "the relative strength of an individual's identification with and involvement in a particular organisation" [5] and consists of affective, normative and continuance commitment [23]. It is particularly important in the IT context as it has a great impact on turnover intention [38]. People with high productivity or fulfilment often perceive themselves as inseparable from their jobs. Consequently, their personal commitment to, and professional engagement with, the organisation for which they work is borne out [24].

Organisational culture and environment are strongly related to affective OC [25, 26]. The person–organisation fit and perceptions of congruence between worker and company sustainably influence employee commitment [27]. The more important values for an employee are reflected in the company the more the employee's commitment rises [28]. Therefore, the active management of embedded organisational culture and values has significant influence on retention.

Another essential point here is how decisions are made within the company. Informal culture and communication, as promoted within the context of an agile system, bind the team members in the form of an organisational family, and therefore, they foster OC [25, 26]. The empowerment to decide regarding work tasks gives an employee the opportunity to contribute to the success of the company, and consequently, it increases their interest in the enterprise's well-being [29, 30].

3 Research Model

The results of the literature review indicate the influence of agile ISD on affective OC beyond agile team members. In our study, we use our research model to investigate the impact of APM practices on affective OC. We hypothesise that APM practices influence affective OC in two ways. First, the impact happens directly through the embedding of positive, agile values into the organisation's processes, thereby affecting the person–organisation fit (H1). Second, the impact of APM practices occurs indirectly consequent to the change in leadership culture by affecting JA (H2a+b) and SS (H3a+b). **Figure 1** illustrates the research model.

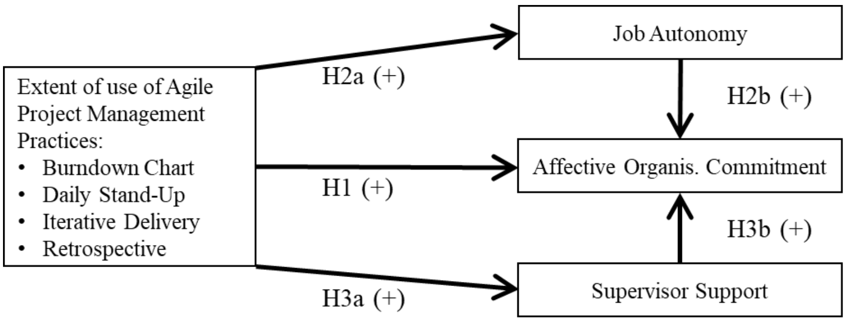


Figure 1. Research model

3.1 Direct Influence of Agile Practices on Affective OC

By promoting agile ISD within the company, the enterprise advocates positive agile values such as collaboration or customer involvement [10]. Retrospectives drive the team to constant reflection and at the same time improve the bond within the team. [31]. The principle of iterative delivery helps team members not only to see individual work packages, but rather to gain an overview of the entire development process [8]. Thus, an agile employee experiences a change at work under the umbrella of attractive and people-oriented values [20].

Agile values and principles play an essential role for agile teams. Williams [32] show that the majority of agile teams (65%) is not only committed to implementing agile practices, but rather to pursuing agile values. We suggest that this integration of agile values within organisational values influences the ways in which employees perceive their organisation. The advocacy of agile values in the company leads to an increased value congruence between employee and company, which implies an increased level of perceived person–organisation fit and supports identification with the enterprise [7, 33] , as well as OC [28]. Thus, we hypothesise the following: *H1: The use of APM practices positively influences affective OC of agile ISD team members.*

3.2 Indirect Influence of Agile Practices on Affective OC

JA, which is described as “the degree to which the job provides substantial freedom, independence and discretion in scheduling the work and in determining the procedures to be used in carrying it out” [34], has increased in the past few years within the IT sector . JA decentralises decision-making to those who actually carry out the work, and thus, it provides the team and the individuals with the flexibility to react faster to changes and unforeseen circumstances [15].

Self-organisation occurs at both the team and individual levels and is a decisive principle of agile practices [11, 13]. In accordance to [8] we suggest that agile methods support JA by combining the doing, the planning, and the controlling of the software development activities and providing the team with the possibility to manage a personal client relationship [8]. Furthermore, agility trusts in the self-coordination of teams and follows a shared decision-making approach [21]. Contrary to traditional software development, where managers delegate, coordinate and supervise work, agile teams are encouraged to organise work independently and make decisions jointly. This principle is, for example, followed by the agile practice of iterative delivery, which encourages team-based estimations regarding workload [35]. Thus, we hypothesise the following: *H2a: The use of APM practices positively influences JA of agile ISD team members.*

JA, as one dimension of intrinsic rewards, positively influences self-disciplined and commitment-driven behaviours. These behaviours are expressed by the presence of hard work, voluntary initiative and the support of organisational objectives [36]. Workers who are self-responsible for the organisation and the coordination of their work are more closely associated with the company.

Furthermore, employees who are challenged with self-organisation at work experience greater well-being and job satisfaction. This is because JA lowers stress, increases motivation and enhances work engagement [8]. Self-organised employees develop a personal interest in the well-being of the company, and consequently, commitment rises and turnover drops [29, 30]. Thus, we hypothesise the following: *H2b: JA positively influences affective OC of agile ISD team members.*

Although agile ISD calls for self-organisation, it does not mean that leadership has become obsolete. To maintain a functioning team, organisational guidance and particularly the support of supervisors are still needed [11, 12, 37]. Agile development requires a new, changed image of managers. Instead of trite adherence to hierarchy and micromanagement, agile ISD calls for management styles that foster the collaborative self-management of teams [37].

Studies already show the positive effects of this new form of leadership. Windeler, Maruping and Venkatesh [38] provide evidence by investigating 73 ISD teams in which empowering leadership contributes to a reduction in role ambiguity, role conflict and stress, and thus, the authors of that study determine that this management style is a suitable measure against technical risk factors [38]. Tyssen, Wald and Heidenreich [39] show that task and people-oriented leadership behaviours in temporary organisations, such as IT projects, are aimed above all at follower commitment in projects, thus having indirect effects on project success [39].

We consider this shift in management styles as beneficial and claim that the change to self-organisation in agile teams does not compensate the need for organisational support; rather, it enhances the perception of SS. By carrying out retrospectives, the team is able to detect problems more quickly and forward them to their supervisor. The use of burndown charts also allows ongoing and rapid feedback from the manager. Thus, we hypothesise: *H3a: The use of APM practices positively influences perceived SS of agile ISD team members.*

Organisational support is one essential key driver of affective OC [40]. By fulfilling socioemotional needs as affiliation and fostering the norm of caring, employees feel the obligation to improve performance and care about the organisation's welfare [41]. Consequently, employees perceive a higher level of satisfaction with their jobs and are more committed to their organisation.

SS, as one dimension of organisational support, reflects the connection between organisational support and affective OC [14, 41]. On the basis of the perceived identification of a supervisor with their organisation, managers act as organisational agents. Therefore, supervisors who appear to be valued by their organisation can highly influence the perceived organisational support of employees [14]. Consequently, we claim that the link between organisational support and commitment can also be transferred to the sub-dimension of SS. Due to flat hierarchies and close cooperation in agile ISD, we suggest that this effect might even be intensified and therefore, we hypothesise the following: *H3b: SS positively influences affective OC of agile ISD team members.*

4 Research Method

4.1 Study Design

Our study was conducted in cooperation with a German company operating in the automotive industry. The company employs approximately 130,000 workers, 4,500 of whom can be allocated to the IT function. The IT organisation strives for a holistic, agile approach, and therefore, it introduced company-wide agile practices in 2016.

We decided to perform the study with the aid of structural equation modelling (SEM). Because of the unestablished character of the theory and the occurrence of 2^{nd} order constructs within the model, we chose Partial Least Squares Structural Equation Modelling (PLS-SEM) over Covariance Based Squares Structural Equation Modelling (CB-SEM) and selected SmartPLS as the software tool [42, 43]. As a foundation for the procedure, we used the instructions of [43].

4.2 Participants and Data Collection Procedure

Due to restrictions, it was not possible to survey the entire IT department of the cooperating company and therefore we had to limit to a smaller target group. To capture a wide range of agile approaches, we intended to reach employees engaged in software engineering, with various degrees of agile experience and with different roles within the teams. On the basis of these parameters, we sent the survey to 380 potential candidates from different departments that use agile methods.

The survey achieved 172 responses, representing a response rate of 45%. The company in which the study was carried out works in close cooperation with external service providers because of which the company employs many people at managerial positions internally. It is, therefore, not surprising that around half of all respondents (52%) stated that they hold project managerial positions, such as Scrum Master Product Owner. The rest of the respondents were Business/System Analysts (15%), Software Developers (9%) and those in other positions (12%). In addition, 12% of all participants belonged to senior management. Most of the respondents had started to use agile methodology within the last 1.5 years (75%), while 6% had medium agile experience (1.5–3 years) and 12% had used agile practices for more than 3 years.

4.3 Measures

For the study, only established measures published in prior research with good quality criteria were chosen. To measure the extent of APM practices, we oriented on [8]. In order to investigate the most common agile practices we conducted a pre-survey within the company with 15 selected representatives from different divisions, with various roles and different degree of agile experience. The pre-survey revealed that iterative delivery (ID), daily stand-up (SU), retrospective (RE) and burndown chart (BD) were the most applied methods in the cooperating company (80 % of all respondents used ID and SU, 70 % used RE and 50% used BD), which is why we decided to integrate these practices in our survey. To assess JA, we used the scale from the job diagnostic survey by [34]. SS was measured using the five-item scale of [44]. For the assessment of affective OC, we used a scale from [45]. All items asked for the participants' agreement on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree).

4.4 Model Analysis

We built the structural equation model based on our research model. In accordance with [8], we defined the construct "APM Practices" as a reflective–formative 2nd-order construct, compounding burndown chart, iterative delivery, retrospective and daily stand-up. The other scales were modelled as reflective as proposed by the respective authors.

Following [46] and [47], we used a multilevel approach for assessing the structural equation models in order to analyse the added effects of agile practices on the suggested model. First, we modelled the direct effects of JA and perceived SS on the dependent variable affective OC (Model 1). Collectively, the manifest variables explained 24.2% of the variance of affective OC. When we included the agile practices (Model 2), the variance of affective OC increased to 26.9%. To test whether this increase was significant, we followed the instructions of [46] and calculated $f^2 = (R^2_{Model2} - R^2_{Model1})/(1 - R^2_{Model2})$. Afterwards, we performed a pseudo F-test ($f2*(n - k - 1)$)[1] [48]. The results indicated a significant change in R^2 ($F = 4.65$, df: 1, 130), which lets us conclude that the construction of the model is valid.

Subsequently, we evaluated the measurement model by conducting tests for internal consistency reliability, convergent validity and discriminant validity. The results of Cronbach's alpha, composite reliability and AVE can be found in **Table 1**. Cronbach's alpha of Retrospective, as well as composite reliability of Retrospective and Supervisor Support showed a critical value of > 0.95. However, in order to avoid losing informative value of the scales by deleting items, we decided to keep the scales as suggested and not to delete any items [43]. In addition, the outer loadings all exceeded the threshold of 0.7. Finally, the results of the cross-loading analysis demonstrated that all items loaded higher on the intended construct than they did on all other constructs, thus implying the construct's discriminant validity [49].

[1] * n = sample size, k = number of independent variables

Table 1: Reliability scores and AVE

		Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
BD	Burndown Chart	.874	.923	.799
ID	Iterative Delivery	.835	.901	.752
RE	Retrospective	.964	.976	.933
SU	Daily Stand-Up	.891	.933	.822
JA	Job Autonomy	.904	.938	.834
SS	Supervisor Support	.947	.960	.827
OC	Organis. Commitment	.820	.880	.648

To examine the validity of our formative 2nd-order construct, we followed the guidelines of [43, 45]. The assessment demonstrated that all 1st-order constructs significantly loaded on the 2nd-order construct. We, therefore, conclude that the modelling of our 2nd-order construct is valid.

Finally, we investigated collinearity, the Q^2 and the R^2 values, to provide evidence for the validity of the structural model. The inner VIF values were all clearly beneath the critical value of 5 and left no indication for a strong correlation between the predictor constructs [43]. Likewise, all Q^2 values, resulting from the blindfolding procedure, exceeded the value of 0, and thus, they provided evidence for the predictive relevance of the model [43]. R^2 of affective OC (0.269) implied a low effect [50].

5 Results

The path coefficient analysis provided evidence for the positive impact of the use of agile practices on affective OC. First, the results show that agile practices significantly and directly influence affective OC (0.201, $p < 0.01$).

This indicates that agile team members are more committed to their organisation. Thus, H1 is fully supported. Furthermore, we found confirmation for indirect links from APM practices to affective OC. JA was significantly influenced by the use of APM practices (0.262, $p < 0.01$), and furthermore, it (JA) impacted affective OC (0.212, $p < 0.05$). These results imply that agile teams perceive more autonomy in their work processes, which consequently leads to a higher level of commitment. Therefore, both Hypotheses 2a and 2b can be supported. In addition, the analysis demonstrated the significant influence of agile practices on SS (0.217, $p < 0.01$) as well as that of SS on affective OC (0.282, $p < 0.01$). Agile team members accordingly feel more supported by their direct supervisors; consequently, affective OC rises. Thus, Hypotheses 3a and 3b are fully supported. The results of the path analysis can be found in **Figure 2**.

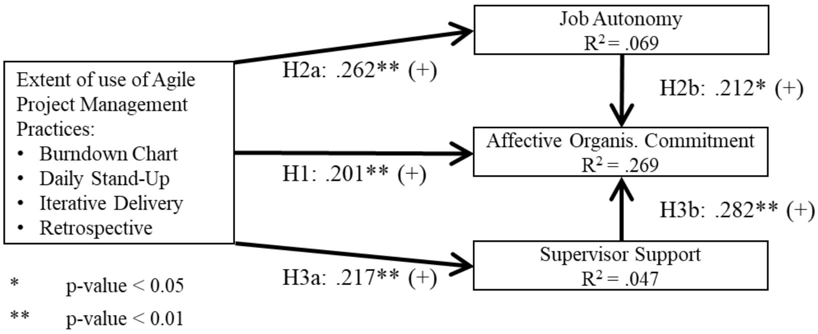


Figure 2. Path analysis results

6 Discussion

The findings of our study are in line with previous research [8, 35]. As with previous studies, we provide evidence for the existing conclusion that the extent of use of agile ISD practices positively predicts JA in agile ISD teams [8]. Furthermore, we reveal that team members of agile ISD teams gain high levels of support and empowerment from their supervisors [35].

We extend existing knowledge and provide insights that go beyond the results of previous research. When focusing on agile ISD practices, prior studies often investigated a particular practice, such as extreme programming or assessed agility, as a concept rather than investigating the extent of its use [51]. Most previous studies only investigated perceptions of work such as job satisfaction [8] and motivation [52]. The present study theorised and validated the effect of the use of agile ISD practices on JA, SS and affective OC. The results show that agile ISD practices have a direct effect on affective OC. In addition, affective OC is indirectly influenced by JA and SS. These additional insights provide several theoretical and practical contributions, which we explain in the following.

6.1 Theoretical Implications

The current research focuses on assessing the motivating and satisfying potential of agile ISD practices [8, 52]. Our theoretically conceptualized and validated research model is the first step in research on agile ISD to consider affective OC as a direct and indirect outcome variable and investigate how JA and SS predict affective OC. Thus, the contribution of our paper to existing literature is threefold.

First, our study contributes by theorizing and validating that agile ISD practices positively effects team members' affective OC. Analyzing the data with structural equation modeling, we show that there is a direct positive relationship between agile practices and affective OC in agile ISD teams. Thus, our paper extends existing knowledge on the effect of agile ISD practices on job perceptions of agile team members. The relationship between agile practices and team motivation was investigated by [52]. As well, Tripp, Riemenschneider and Thatcher [8] found that team members are more satisfied if agile ISD practices are used by showed that particular agile practices are related to job characteristics, such as skill variety and feedback. Therefore, we contribute to the agile ISD literature by indicating that in addition to increased satisfaction and motivation, team members are also higher committed to their organisation. By investigating affective OC in agile ISD teams, we also contribute to research on turnover of IS professionals in general [30]. Van Scotter and Motowidlo [36] found that affective OC is a valid predictor of employees' turnover intention. Existing studies often investigated teams using particular practices (such as Pair Programming [53]). This approach has gained critique as it does not assess the amount of use of agile practices, but rather the "high level concept of use of that method" [51]. Our study extends existing knowledge on agile practices as we measured the extent of use of agile practices, as suggested by [51]. In summary, we contribute to literature on agile by demonstrating that the extent of use of agile practices positively affects team members' affective OC.

Second, prior literature indicates that the extent of use of agile practices is positively related to team members' JA [8]. Our study replicates the findings in a conceptual manner [54] and shows a significantly positive relationship between the use of agile practices and team members' JA. As well, we found that JA positively affects affective OC in agile ISD teams. Based on previous literature indicating that self-organisation is a decisive principle of agile practices [11, 13], we theorized that the amount of agile practices positively predicts autonomy in agile ISD teams. The use of agile practices decentralizes decision-making to those who actually carry out the work, providing the team and the individuals with the flexibility to faster react to changes and unforeseen circumstances [15]. Thus, we argue that this principle of self-organisation is fostered when agile practices are used. As well, we theorized based on literature indicating that self-organised employees develop personal interest in the well-being of the company and, as a consequence, higher commitment [29], that JA predicts affective OC in agile ISD teams. Employees who are self-responsible for the organisation and the coordination of their work are more closely associated with the company. These theoretically derived effects have been proven within our study. Therefore, our study contributes to research as we replicate the results of [8] by indicating that the use of agile practices directly influences team members' JA. In addition, our results extend research focusing on affective OC [29] by revealing JA as a predictor of affective OC in agile ISD teams. Furthermore, the findings of our study extend prior literature on self-organisation and turnover such as [30] by providing evidence that JA significantly predicts affective OC in agile ISD teams. Moreover, our results also extend existing general literature on self-organisation of agile ISD teams such as [12], [13] and [22] by revealing that there is not only a direct effect of self-organisation on affective OC, but of JA on affective OC as well. In summary, we contribute to literature on self-organisation of agile ISD teams by providing evidence for a positive effect of agile practices on JA as well as a positive relationship between JA and affective OC.

Third, our study reveals that the use of agile ISD practices significantly influences team members' SS and this, in turn, predicts their affective OC. Hence, with an increasing amount of use of agile practices, team members gain more support from their supervisor. As a result, when perceiving themselves as being highly supported by their leader, team members feel more committed to the organisation. The present paper contributes to existing literature such as [55], who state that empowering leadership is crucial for project performance in agile ISD teams, by demonstrating that agile practices significantly affect team members' perceived support by their supervisor. agile Moreover, we contribute to existing knowledge such as [38], who find that empowering leadership is related to lower role ambiguity, role conflict and stress in software development, by showing that perceived SS is also positively related to affective OC. By fulfilling socioemotional needs as affiliation and fostering the norm of caring, employees feel the obligation to increase performance and care about the organisation's welfare [41]. In consequence, they are more committed to their organisation. Additionally, we extent literature such as [39], who found that people-oriented leadership behaviours in IT projects are aimed at follower commitment, by revealing that SS positively predicts affective OC. Summarizing, the present paper contributes to agile literature and psychological commitment literature by revealing that agile practices positively affect SS, which in turn is positively related to affective OC.

6.2 Practical Implications

In addition to theoretical derivations, our study also provides insights for practitioners. The results of the study demonstrate that agile ISD significantly influences not only work processes but also the well-being of IT professionals. By means of APM methods, it is possible to influence the perception of employees towards their company. Through striving for agile principles, an employee finds a common basis with their organisation and feels more connected to their company.

In addition, the results highlight the impact of leadership in empowered teams. The study proves that autonomy in agile teams by no means makes leadership obsolete but rather puts it at the centre of the scheme of things. The results demonstrate that JA and SS are of equal importance for an employee. This indicates the necessary shift in the current leadership culture. Supervisors must position themselves in a supportive and protective function to further encourage the self-organisation of their teams. This concept enhances the well-being of individuals and retains employees for the long term.

6.3 Limitations and Future Work

We want to emphasise the importance of attachment in the IT environment, thus encouraging future work regarding affective OC in the IT context. In this regard, we want to draw attention to our limitations to correctly classify the interpretation of the results and simultaneously create space for future research. First, we point out the importance of an enriched replication of this study. Hypothesis 1 is based on the argument that the introduction of agile values leads to an increase in person-organisation fit, which consequently leads to a rise in organisational commitment. This reasoning was derived from research by [7] and [28]. Unfortunately, we were not able to measure the person-organisation fit aspect in our study. Furthermore, we would like to point out that the scales used only allow a measurement of the application of agile practices, but at the same time do not indicate that agile values are actually adapted in the company. We consider the investigation of the influence of agile practices on person-organisation fit as an important field of research, which is why we recommend to inspect this in future research. Second, our study was explorative in character, and therefore, it includes modest R^2 values. Because of the explorative character of this study, these results are comprehensible. We compared the variance with other explorative research in the area of affective OC and found similar results [e.g. 33, 56] because of which we think that our study brings essential contribution to the current agility research. In future, a replication of the study with a bigger sample size could lead to results with higher variance. Finally, we would also like to point out the possibility of reverse effects in the model. As OC leads to positive behavior [57], we recognize that the extent of agile practices may also be affected by the appearance of affective OC. For future research we therefore also recommend an investigation of the interaction of the presented independent and dependent variables.

7 Conclusion

Our paper provides insights into the influence of APM practices on affective OC. Based on structural equation modelling, our results demonstrate (1) the direct impact of agile ISD on affective OC via affecting values within the organisation and (2) the indirect impact of agile ISD on affective OC via following the agile principle of self-organisation and the consequent fostering of JA and SS. Our study, therefore, sheds light on the effects of agile practices on affective OC, while emphasising the importance of empowered teams and their relationship with leadership in the IT context.

References

1. Gartner, Inc., https://www.gartner.com/newsroom/id/3871063 (Accessed: 29.08.2018)
2. Bitkom, https://www.bitkom.org/Presse/Presseinformation/Bitkom-Branche-ist-erstmals-groesster-Industrie-Arbeitgeber.html (Accessed: 29.08.2018)
3. Joseph, D., Ng, K.Y., Koh, C., Ang, S.: Turnover of information technology professionals: A narrative review, meta-analytic structural equation modeling, and model development. MIS Quarterly 31, 547-577 (2007)
4. LinkedIn, https://business.linkedin.com/talent-solutions/blog/trends-and-research/2018/the-3-industries-with-the-highest-turnover-rates (Accessed: 29.08.2018)
5. Mowday, R.T., Steers, R.M., Porter, L.W.: The measurement of organizational commitment. Journal of Vocational Behavior 14, 224–247 (1979)
6. Porter, L.W., Steers, R.M., Mowday, R.T., Boulian, P.V.: Organizational commitment, job satisfaction, and turnover among psychiatric technicians. Journal of applied psychology 59, 603 (1974)
7. Finegan, J.E.: The impact of person and organizational values on organizational commitment. Journal of Occupational & Organizational Psychology 73, 149-169 (2000)
8. Tripp, J.F., Riemenschneider, C., Thatcher, J.B.: Job Satisfaction in Agile Development Teams: Agile Development as Work Redesign. Journal of the Association for Information Systems 17, 267–307 (2016)

9. Tessem, B., Maurer, F.: Job satisfaction and motivation in a large agile team. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 4536 LNCS, pp. 54-61 (2007)
10. Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R., Mallor, S., Shwaber, K., Sutherland, J.: Manifesto for Agile Software Development. (2001)
11. Cockburn, A., Highsmith, J.: Agile software development, the people factor. Computer 34, 131–133 (2001)
12. Hoda, R., Noble, J., Marshall, S.: Self-Organizing Roles on Agile Software Development Teams. IEEE Transactions on Software Engineering 39, 422–444 (2013)
13. Moe, N., Dingsøyr, T., Dybå, T.: Understanding Self-Organizing Teams in Agile Software Development. Proceedings of the Australian Software Engineering Conference, ASWEC 76-85 (2008)
14. Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I.L., Rhoades, L.: Perceived supervisor support: Contributions to perceived organizational support and employee retention. Journal of Applied Psychology 87, 565–573 (2002)
15. Lee, G., Xia, W.: Toward agile: an integrated analysis of quantitative and qualitative field data on software development agility. Mis Quarterly 34, 87-114 (2010)
16. Conboy, K.: Agility from first principles: Reconstructing the concept of agility in information systems development. Information systems research 20, 329–354 (2009)
17. Erickson, J., Lyytinen, K., Keng, S.: Agile Modeling, Agile Software Development, and Extreme Programming: The State of Research. Journal of Database Management 16, 88–100 (2005)
18. Pflügler, C., Wiesche, M., Krcmar, H.: Subgroups in Agile and Traditional IT Project Teams. 51st Hawaii International Conference on System Sciences (HICSS), (2018)
19. Przybilla, L., Wiesche, M., Krcmar, H.: The Influence of Agile Practices on Performance in Software Engineering Teams: A Subgroup Perspective. Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research, pp. 33-40. ACM (2018)
20. Lalsing, V.: People Factors in Agile Software Development and Project Management. International Journal of Software Engineering & Applications 3, 117-137 (2012)
21. Moe, N.B., Aurum, A., Dybå, T.: Challenges of shared decision-making: A multiple case study of agile software development. Information & Software Technology 54, 853–865 (2012)
22. Shastri, Y., Hoda, R., Amor, R.: Understanding the roles of the manager in agile project management. Proceedings of the 10th Innovations in Software Engineering Conference, pp. 45-55. ACM (2017)
23. Meyer, J.P., Allen, N.J.: A three-component conceptualization of organizational commitment. Human resource management review 1, 61–89 (1991)
24. Chalofsky, N., Krishna, V.: Meaningfulness, Commitment, and Engagement: The Intersection of a Deeper Level of Intrinsic Motivation. Advances in Developing Human Resources 11, 189–203 (2009)
25. Paul, A.K., Anantharaman, R.N.: Influence of HRM practices on organizational commitment: A study among software professionals in India. Human Resource Development Quarterly 15, 77–88 (2004)
26. Messner, W.: Effect of organizational culture on employee commitment in the Indian IT services sourcing industry. Journal of Indian Business Research 5, 76-100 (2013)
27. Van Vianen, A.E.: Person-Organization Fit: The Match Between Newcomers'and Recruiters'Preferences For Organizational Cultures. Personnel psychology 53, 113-149 (2000)

28. O'Reilly Iii, C.A., Chatman, J., Caldwell, D.F.: People and Organizational Culture: A Profile Comparison Approach to Assessing Person-Organization Fit. Academy of Management Journal 34, 487-516 (1991)
29. Aube, C., Rousseau, V., Morin, E.M.: Perceived organizational support and organizational commitment: The moderating effect of locus of control and work autonomy. Journal of managerial Psychology 22, 479-495 (2007)
30. McKnight, D.H., Phillips, B., Hardgrave, B.C.: Which reduces IT turnover intention the most: Workplace characteristics or job characteristics? Information & Management 46, 167-174 (2009)
31. McHugh, O., Conboy, K., Lang, M.: Agile practices: The impact on trust in software project teams. IEEE Software 29, 71-76 (2012)
32. Williams, L.: What agile teams think of agile principles. Communications of the ACM 55, 71-76 (2012)
33. Venkatesh, V., Windeler, J.B., Bartol, K.M., Williamson, I.O.: Person–organization and person–job fit perceptions of new IT employees: Work outcomes and gender differences. Management Information Systems Quarterly 41, 525-558 (2017)
34. Hackman, J.R., Oldham, G.R.: Development of the job diagnostic survey. Journal of Applied Psychology 60, 159-170 (1975)
35. Tessem, B.: Individual empowerment of agile and non-agile software developers in small teams. Information and Software Technology 56, 873–889 (2014)
36. Van Scotter, J.R., Motowidlo, S.J.: Interpersonal facilitation and job dedication as separate facets of contextual performance. Journal of applied psychology 81, 525-531 (1996)
37. Moe, N.B., Dingsøyr, T., Dybå, T.: Overcoming Barriers to Self-Management in Software Teams. IEEE Software 26, 20-26 (2009)
38. Windeler, J.B., Maruping, L., Venkatesh, V.: Technical systems development risk factors: The role of empowering leadership in lowering developers' stress. Information Systems Research 28, 775-796 (2017)
39. Tyssen, A.K., Wald, A., Heidenreich, S.: Leadership in the context of temporary organizations: A study on the effects of transactional and transformational leadership on followers' commitment in projects. Journal of Leadership & Organizational Studies 21, 376-393 (2014)
40. Loi, R., Hang-Yue, N., Foley, S.: Linking employees' justice perceptions to organizational commitment and intention to leave: The mediating role of perceived organizational support. Journal of Occupational and Organizational Psychology 79, 101-120 (2006)
41. Rhoades, L., Eisenberger, R.: Perceived organizational support: a review of the literature. Journal of applied psychology 87, 698-714 (2002)
42. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. SmartPLS GmbH
43. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: Partial Least Squares Strukturgleichungsmodellierung (PLS-SEM). Verlag Franz Vahlen GmbH, München (2017)
44. Cousins, R., Mackay, C.J., Clarke, S.D., Kelly, C., Kelly, P.J., McCaig, R.H.: 'Management standards' work-related stress in the UK: Practical development. Work & Stress 18, 113–136 (2004)
45. Sjöberg, A., Sverke, M.: The interactive effect of job involvement and organizational commitment on job turnover revisited: A note on the mediating role of turnover intention. Scandinavian Journal of Psychology 41, 247–252 (2000)
46. Siponen, M., Vance, A.: Neutralization: New insights into the problem of employee information systems security policy violations. MIS Quarterly 487–502 (2010)

47. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 36, 157–178 (2012)
48. Mathieson, K., Peacock, E., Chin, W.W.: Extending the technology acceptance model: The influence of perceived user resources. ACM SIGMIS Database: the DATABASE for Advances in Information Systems 32, 86–112 (2001)
49. Gefen, D., Straub, D.: A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example. Communications of the Association for Information Systems 16, 91-109 (2005)
50. Henseler, J., Ringle, C.M., Sinkovics, R.R.: The use of partial least squares path modeling in international marketing. New challenges to international marketing, pp. 277–319. Emerald Group Publishing Limited (2009)
51. Tripp, J.: Thoughts on Current and Future Research on Agile and Lean: Ensuring Relevance and Rigor. Hawaii International Conference on System Sciences, pp. 5465-5472, Hawaii (2018)
52. McHugh, O., Conboy, K., Lang, M.: Using agile practices to influence motivation within IT project teams. Scandinavian Journal of Information Systems 23, 59-84 (2011)
53. Salge CA de Lima, Berente, N.: Pair Programming vs. Solo Programming: What Do We Know After 15 Years of Research? 49th Hawaii International Conference on System Sciences, pp. 5398-5406 (2016)
54. Dennis, A.R., Valacich, J.S.: A replication manifesto. AIS Transactions on Replication Research 1, 1-4 (2014)
55. Shen, Y., Xu, P.: Leading Agile Teams: An Exploratory Study of Leadership Styles in Agile Software Development. Twenty-first Americas Conference on Information Systems, Puerto Rico (2015)
56. Jaramillo, F., Mulki, J.P., Marshall, G.W.: A meta-analysis of the relationship between organizational commitment and salesperson job performance: 25 years of research. Journal of Business Research 58, 705-714 (2005)
57. O'Reilly III, C., Chatman, J.: Organizational Commitment and Psychological Attachment: The Effects of Compliance, Identification, and Internalization on Prosocial Behavior. Journal of Applied Psychology 71, (1986)

The Complexity Trap – Limits of IT Flexibility for Supporting Organizational Agility in Decentralized Organizations

Tobias Segert[1], Friedrich Holotiuk[1], and Daniel Beimborn[2]

[1] Frankfurt School of Finance & Management, Frankfurt, Germany
{t.segert, f.holotiuk}@fs.de
[2] University of Bamberg, Bamberg, Germany
daniel.beimborn@uni-bamberg.de

Abstract. In times of digitalization, organizations are transforming to become more agile. Decentralizing decision-making to lower levels of the organization allows firms to better detect changes in their environment and to sense new opportunities. Simultaneously, the digital transformation of a firm is increasingly relying on a flexible IT function. This study explores how decentralization of decision-making power increases organizational agility and to what extent this relationship is dependent on IT flexibility. We conducted a quantitative study using two rounds of surveys with 123 respondents from the consulting industry. Based on covariance-based SEM, we find that too much flexibility of the IT function in combination with decentralized decision-rights creates a complexity trap and, thus, decreases organizational agility. These findings carry several theoretical contributions for organizational agility research and shed new light on the role of IT flexibility for digital transformation.

Keywords: Organizational Agility, Decentralization, IT Flexibility, Detecting Capability, Digitalization

1 Introduction

The concept of organizational agility has been discussed as a solution for organizations to prosper in volatile environments. The term *agility* was coined in the Manufacturing and Information Systems (IS) literature to define a firm's ability to seize opportunities due to changes in the environment [1]. Today, this concept is more relevant than ever: digitalization is creating an increasingly complex and unpredictable environment [2]. Digitalization is "a sociotechnical process of applying digitizing techniques to broader social and institutional contexts that render digital technologies infrastructural" [4 p. 2]. Among others, it changes market boundaries, substitutes physical labor with new technology, eliminates cognitive labor by increased application of software algorithms, and replaces physical products and services with new digital offerings [2].

Trying to compete in such a fast-changing environment is a complex task and organizations struggle to find an efficient organizational design that promotes their

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

adaptability to those unforeseeable changes [4]. One solution to transform the organization is to delegate responsibilities to lower levels of the organization to increase the employees' ability to respond to changes. Research has shown that increased decentralization has been beneficial towards organizations' ability to respond to change. However, extant theory is scarce in explaining why this relationship exists [5].

Today, firms' operations are highly dependent on IT [3]. In times of digitalization, the importance of digital options has been stressed as an enabler of organizational agility [6], [7]. In this context, Byrd and Turner [8] emphasize the supporting role of the IT function for organizations. Even though the enabling roles of both IT flexibility and decentralization are separately well recognized in the literature, the question remains how they are interrelated with regard to organizational agility. Hence, we derive the following research question: *How does the decentralization of decision-making rights complement with IT flexibility in enabling organizational agility?*

To answer this question, our research uses a structured survey, collecting data at two points in time. We develop a conceptual model based on constructs from extant research. We draw on a sample from the professional service industry and analyze our hypotheses using covariance-based SEM. Our findings uncover the limits of IT flexibility as firms are experiencing a *complexitiy trap* when high decentralization and high IT flexibility interfere. The remainder is structured as follows: section 2 introduces the conceptual model and hypotheses. Afterwards, the sample and measurement development are explained in section 3. Section 4 presents the findings based on the tests of the hypotheses. Finally, we discuss the main implications and conclude the paper.

2 Model Development

Our theoretical model (Figure 1) is derived from extant literature.

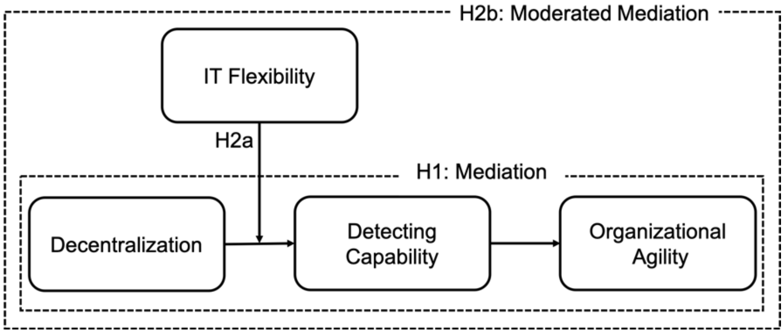


Figure 1. Conceptual Model

First, we conceptualize how decentralization of decision rights enhances agility by proposing detecting capability as a mediator. Second, we examine to what extent this relationship is conditional on IT flexibility. Understanding these mechanisms can help designing organizations that rely on the flexibility of IT to face the challenges of digitalization. A summary of all constructs with definitions can be found in Table 1.

Within the IS literature, agility had first been used in the software development domain. Later, scholars also applied the concept of agility to the entire organization. They offer new theoretical arguments for how the organizations' ability to thrive in a digitalized world is dependent on IT capabilities. Scholars point out the difference between agility and flexibility [9], responsiveness [10], dynamic capabilities [9] and strategic flexibility [11]. Zhang and Sharifi [12] define the concept of agility as the ability to adequately respond to anticipated or unexpected changes in the environment (like the advent of new market players or digital technologies) and to see change as an opportunity. Additionally, scholars have noted the importance of sensing change for the concept of organizational agility [6], [13].

Table 1. Summary of key constructs

Construct	Definition	Reference
Decentralization	"The extent to which power over decision making in the organization is dispersed among its members" with a focus on vertical decentralization which "refers to the extent to which formal decision-making power is "delegated" down to the chain of line authority."	[15 p. 326]
IT Flexibility	Ability of the IT function to provide and support the business unit with ex-post adjustments to IT applications.	[15]
Detecting Capability	Firm-wide capability to proactively detect sudden changes in the business environment.	[6]
Organizational Agility	"Firm-wide capability to deal with changes that often arise unexpectedly in business environments via rapid and innovative responses that exploit changes as opportunities to grow and prosper"	[8 p. 933]

2.1 Detecting Capability as a Mediator of the Relationship between Decentralization and Organizational Agility

Based on existing research on agility, we have identified two established research streams as foundation for our model: one stream focuses on the importance of detecting capabilities for organizational agility, and the other, stemming from organizational design, looks at the role of decentralized decision making for detecting capabilities.

Detecting Capability and Organizational Agility

First, we follow the initial definition of organizational agility, defined as a "firm-wide capability to deal with changes that often arise unexpectedly in business environments via rapid and innovative responses that exploit changes as opportunities to grow and prosper" [8 p. 933]. The detecting capability, however, is defined as the firm-wide capability to proactively detect sudden changes in the business environment [6].

We argue that the concept of detecting environmental change and the concept of responding to these changes are discriminant in how they are facilitated in organizations. The ability to detect change is formed, e.g., by the capability to track competitors' actions, to identify changes in the customer's demand, or to recognize shifts in the economy [11]. By contrast, the capability to react to change encompasses the ability to make rapid changes to organizations' products, IT applications, supply chains, resource allocation, or operations [7], [11]. The rationale is that the constant screening of the environment and the detection of possible changes enables organizations to quickly act and capitalize on this change. We conclude that organizations which are aware of threats and opportunities posed in its environment are better able to react to those changes. Consequently, organizations with superior detecting capability are better prepared to react quickly to changes.

Decentralization and Detecting Capability

According to Winkler and Wessel [16], decision rights are concerning different research themes like IT governance and firm structure. We concentrate on the latter theme, which emphasizes the influence of information systems on firm structures [16]. Following Mintzberg [14], there are two distinct types of decentralization on the organizational level: horizontal and vertical decentralization. We focus on vertical decentralization, which describes "the extent to which formal decision making power is 'delegated' down to the chain of line authority" [15 p. 326]. The delegation of decision rights to lower levels of the hierarchy encourages lateral and vertical communication across employees [5] and increases employees' satisfaction and motivation [17]. As Hage and Dewar [18] point out, centralized decision-making power in the hands of a few impedes the flow of creative thoughts across various members within the organizations and blocks innovative ideas. The rationale is that innovative ideas pose threats to the existing distribution of power in organizations. In centralized organizations, innovative ideas must pass several managers along the chain of command before receiving approval or resource support and are, therefore, more likely to be rejected [19].

Therefore, organizations with a decentralized structure, having many empowered employees on lower levels of their hierarchy, show a better flow of ideas concerning the opportunities and threats they are facing. Employees in decentralized structures are also more motivated [17] to screen and understand the complex changes of the now digitalized environment that influence their work. We conclude, organizations with a high level of decentralization can better detect and understand changes in their environment, e.g., in the form of new IT innovation, or changing customer demands.

Decentralization on Organizational Agility *via* Detecting Capability

Following the extant literature, the positive relationships of decentralization on detecting capability and of detecting capability on organizational agility are well-established. Still, there are different findings regarding the relationship between decentralization and organizational agility. On the one hand, literature suggests that a high decentralization of decision-making power is associated with high organizational agility [20]. Delegating decision-making rights down the chain of command is believed to increase the speed in decision making allowing for faster reaction to changes. On the other hand, scholars point out that decentralized decision-making limits organizations' ability to respond to immediate threats [21]. If every subunit in the organization has the authority to react independently to external changes, the organization might act in many different and possibly even contrary ways. Hence, decentralization requires additional coordination efforts to manage the independent subunits or employees to overcome increased complexity [22].

Based on the mixed findings regarding the association between decentralization and organizational agility [21], and based on the fact that the relationships between detecting capability and organizational agility [11] and between decentralization and detecting capability [18] are well established, we propose that detecting capability acts as a mediating variable. Thus, detecting capability explains the effect of decentralization on organizational agility. This means that the effect of decentralization on organizational agility is due to the detecting capability. The vertical decentralization of decision rights enhances organizations' ability to detect changes posed in the environment, which increases organizational agility. Accordingly, we suggest that decentralization is not directly increasing organizational agility. In fact, decentralization only indirectly enhances organizational agility by increasing organizations' capability to detect change. We hypothesize a mediating effect:

The effect of decentralization on organizational agility is mediated by detecting capability. (H1)

2.2 The Moderating Effect of IT Flexibility

In times of digitalization, the role of IT units is heavily discussed. We regard IT flexibility as the ability of an IT unit to provide and support the business side with ex-post adjustments to IT applications [15]. This focuses less on a technical understanding (such as: modular, state-of-the-art IT infrastructures), but rather on an organizational understanding (such as: defined responsibilities, IT service processes, and sufficient resources) of IT flexibility and matches the business side's perception of IT flexibility.

The literature holds different perspectives which offer opposite indications on the effect of IT flexibility. Whereas positive effects of IT flexibility are reasoned with more tailored IT applications supporting diverse business units, negative effects of IT flexibility are reasoned with increased complexity caused by too many and unstandardized IT applications which cripple the IT function [23]. While Overby et al. [11] and Sambamurthy et al. [6] propose a direct effect of digital options on organizational agility, we follow the argumentation of Tallon and Pinsonneault [13] that IT flexibility can also act as a moderator variable in agility research. In the context of our study, using IT flexibility as a moderator allows us to conceptualize that organizations with similar degrees of decentralization could have different levels of detecting capability depending on the flexibility of their IT function. This yields new insights into whether the effect of decentralized decision-making on the organizations' detecting capability might be stronger in combination with a flexible or inflexible IT.

Currently, we find two opposite perspectives in the literature about how decentralization and IT flexibility complement each other. Some scholars argue for a positive interaction effect between IT governance decentralization and the degree of IT architecture modularity when it comes to IT agility [24]. The rationale is, that a highly modular IT architecture in combination with a high IT governance decentralization reduces the coordination efforts between different departments and enhances the speed when it comes to decisions regarding the adaptation of new IT applications. However, the idea of decentralizing decision rights in IT governance is distinct to the idea of decentralizing decision rights in the overall firm structure [16]. In addition, our conceptualization of IT flexibility as the ability of the IT function to support the business side with ex-post adjustments to IT applications is different to the conceptualization of IT architecture modularity, which is defined by loosely coupled subsystems that communicate through standardized interfaces [24]. Another argumentation for a positive interaction effect of IT flexibility and decentralized decision-rights is provided by the increased availability of digital options [13], which allows decentralized business units to reap the benefits of tailored IT applications to improve their process of detecting changes.

Contrary to the positive interaction effect between decentralization and IT flexibility, literature suggests a negative interaction effect based on findings about IT silos and shadow IT. In decentralized organizations, employees are empowered to make own decisions. While this is generally associated with positive effects (e.g., satisfaction and motivation), it can be unfavorable when employees are allowed to order many tailored IT solutions through a flexible IT function, because the diversity and number of IT applications in the organization increases complexity [23]. As Tiwana and Konsynski [24] point out, interdependencies of IT applications spanning across the organization pose several threats to IT agility when changes in the existing IT application portfolio are made. Since not all members of the organization have a holistic understanding of the interplay of the organization’s IT applications, they cannot recognize the problems that surface when integrating or adapting new IT applications. Thus, negative effects of IT flexibility could kick in and cause organizations to end up in a *complexity trap,* which reduces the alertness for change like new IT opportunities [24]. For example, when employees on the lower level of the organization are delegated to assess the challenges and opportunities of a new product, highly flexible internal IT offerings will likely lead to employees working with different methods based on different IT applications they deem appropriate (they are authorized to decide what is appropriate). Thus, comparing results of the assessment becomes more complex and causes uncertainty in the organizations’ detecting process [23].

In summary, there are two opposite predictions about how decentralization and IT flexibility complement each other with regard to the detecting capability of an organization. Hence, we derive the following hypothesis with an unknown direction of the moderating effect of IT flexibility:

The relationship between decentralization and detecting capability is moderated by IT flexibility. (H2a)

The argumentation for a moderation effect of IT flexibility on the relationship between decentralization and detecting capability also suggests moderated mediation. Since we argue, that detecting capability mediates the relationship between decentralization and organizational agility (H1), and the direct relationship between decentralization and detecting capability is moderated by IT flexibility (H2a), we also hypothesize that the mediating effect is moderated by IT flexibility. This means that organizations with the same level of decentralized decision-making end up with different levels of organizational agility dependent on the degree of the flexibility of their IT function. This is, as argued above, because the effect of decentralization on detecting capability is not stable, but changes depending on different levels of IT flexibility:

The indirect relationship between decentralization and organizational agility via detecting capability is moderated by IT flexibility. (H2b)

3 Method

To test our hypotheses, we conducted an empirical study. We developed two structured questionnaires. The first survey contained multiple-item scales for the measurement of three constructs (decentralization, IT flexibility, and detecting capability), the second survey measured organizational agility.

3.1 Sample

Data was collected from the consulting industry. The market for standard consulting services has become increasingly competitive, leading to eroding margins [25]. New technology-based services such as data analytics and artificial intelligence applications are becoming increasingly important and change clients' demand. This forces consultancies to compete in markets that were once solely occupied by technology companies [25]. Compared to other industries, the need to stick to certain IT applications is less rigid in the consulting industry. In the manufacturing industry, for example, applications are interfacing with adjacent firms to a much higher degree. Besides the quick change of clients' demand, Software-as-a-Service offerings by large business providers, such as SAP with S/4Hana Cloud, increase technology companies' grip on IT related services [25]. Since agility allows organizations to stay competitive, it is highly important in the professional service industry. We targeted consultants in both higher and lower hierarchical positions in their respective company to answer our survey. Therefore, all hierarchical levels of the organizations are represented: 50% of the respondents are consultants, 30% are managers, and 15% belong to the top management of their consultancy. About one third of the consultancies have an IT focus, whereas one quarter focuses on auditing and another quarter on strategy and process consulting. Therefore, our dataset encompasses a good mix of the different specializations within the consulting industry. About 40% of the consultants that answered the survey have been working in their consultancy for more than four years, so we assume that consultants in our sample have sufficient knowledge to give reasonable answers to our survey.

3.2 Data collection procedure

An invitation to participate in the first online survey round was distributed through various channels. The first survey round was open for completion from June to August in 2017. Only consultants were allowed to answer the survey. The survey was mainly sent out to consultants in Germany. Consultants were assured that all answers would be kept anonymous. The survey was distributed via the authors' personal networks, the university's network, and professional social networks Xing and LinkedIn. Since consultants were also asked to forward the invitation to their colleagues, we cannot report the final number of contacted consultants. For the second survey round, personal links were sent out to those 181 consultants that answered the first survey *and* left their email address for a second survey. In this second survey, we collected data about the dependent variable organizational agility. The survey was open for completion from May to June in 2018. Ultimately, 123 consultants participated in both of our survey rounds.

3.3 Measures

The survey contained multiple-item scales for the measurement of decentralization, IT flexibility, detecting capability and organizational agility. Adaptations to the original questions were made to reflect the consulting industry's specialties and to improve the consultants' understanding of the questions [26]. An overview of the constructs and items can be found in the appendix. The items were developed in English; we followed the standard translation and back-translation procedure to create German equivalent versions of the measures. All items were pre-tested in both languages with potential participants and fellow researchers to ensure content validity. Means, standard deviations, and correlations among the latent variables are presented in Table 2.

Table 2. Constructs measures and quality criteria

	Variable	1	2	3	4	CA	CR	AVE
1	Decentralization	*.684* [a]				.719	.724	.468
2	IT Flexibility	.285*	*.768*			.803	.810	.590
3	Detecting Capability	.623**	.459**	*.771*		.710	.814	.594
4	Organizational Agility	.296*	.376**	.636**	*.737*	.779	.780	.543

[a] Square roots of the AVE reported on the diagonal in italic, N = 123, * $p < .05$ / ** $p < .001$
CA = Cronbach's Alpha, AVE = Average Variance Extracted

Measures for the independent and dependent variables were collected at two different points in time: First, the perceived horizontal decentralization of consultancy was measured with a five-point Likert scale, using a three-item measure from Claver-Cortés et al. [27]. Decentralization in this research is the extent to which a consultant experiences decision rights being distributed to lower levels of the chain-of-command. Consultants rated the perceived flexibility of their IT on a five-point Likert scale, using a three-item measure from Martin et al. [15]. IT flexibility in this research is the extent to which the IT function is able to provide and support ex-post adjustments to IT applications. The detecting capability of organizations was measured with a seven-point Likert scale as this construct is a dependent variable. Based on Ahsan and Ngo-ye [28] and Sambamurthy et al. [6] we derived three items. Detecting capability comprises proactively detecting changes in the business environment. Lastly, organizational agility is defined by Lu and Ramamurthy [7] with two dimensions: market capitalization agility and operational adjustment agility. To maintain a clear conceptual separation between the idea of detecting and reacting to change, we focus on operational adjustment agility. This dimension emphasizes the reactivity of operational activities as an enabler of innovative initiatives in times of change [7] and, thus, represents the original idea of organizational agility by focusing on the organizations' ability to react to changes [12], [13]. Additionally, the measurement model would not structurally change when using a first-order construct for organizational agility with two dimensions (market capitalization agility and operational adjustment agility). We measure organizational agility using the three items of the operational adjustment dimension by Lu and Ramamurthy [7].

3.4 Testing for Biases

Although a time-lag design (surveys at two points in time) mitigates the concern of common method variance, several additional procedural and statistical remedies as suggested by Podsakoff et al. [26] were taken into consideration. Participants were assured anonymity and confidentiality of their responses to limit concerns about evaluation apprehension and social desirability. Buffer items, not relevant for this research model, were included in the survey. We also checked for nonresponse bias. The first 30% of the responses were compared with the last 30% of the responses for both survey rounds. The independent sample t-test found no statistically significant difference ($p < .05$), indicating that nonresponse bias is not an issue in our study. Finally, comparing the responses for the German and English version of the survey did not result in any statistically significant difference ($p < .05$).

4 Results

4.1 Validation of Measurement Model

We run a confirmatory factor analysis with Mplus (Version 8.1) to test whether the collected data reflects the hypothesized measurement model [30] and to establish convergent and discriminant validity of our measures. Following Hu and Bentler [31], the overall model fit was very good (χ^2 (48) = 49.34, CFI = .997, TLI = .996, SRMR = .038, RMSEA = .015). All of our fit indicators are above the threshold of .95 for the CFI and TLI value and below the thresholds of .08 and .06 for the SRMR and RMSEA value, suggesting a good model fit [31]. Each item loaded on its appropriate factor, all standardized factor loadings were significant ($p < .000$) and larger than .59. The correlation among all latent constructs were all substantially smaller than 1, and the square root of the AVE for each construct was larger than its highest correlation with any other construct (see Table 2), indicating that our constructs are sufficiently discriminant.

4.2 The Mediating Role of Detecting Capability

We applied covariance-based structural equations modeling (CBSEM) to examine the relationships among the latent variables. Following Cheung and Lau [32], we used a three-step approach for testing moderated mediation with latent moderated structural equations (LMS). This procedure produces more accurate parameter estimates and confidence intervals than commonly used regression-based approaches, which rely on observed variables and do not correct for measurement errors [32]. We used Mplus with maximum likelihood estimation to examine the moderated mediation model.

First, because the LMS approach does not provide the usual fit indices, we assessed the overall model fit without the interaction (moderation) term following Cheung and Lau [32]. The fit indices indicate that the suggested model fits the data very well (χ^2(49) = 50.19, CFI = .998, TLI = .997, SRMR = .039, RMSEA = .014).

In a second step, we evaluated the whole model (Figure 2) including the interaction between decentralization and IT flexibility. 2,000 bootstrap samples were generated to examine the model. Standardized path coefficients (β) are reported in Figure 2. Our model explains 44 percent of the variance in detecting capability ($R^2 = .435$) and 49 percent of the variance in reacting agility ($R^2 = .489$).

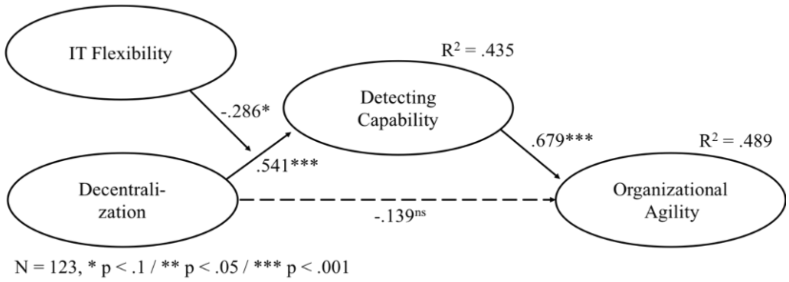


Figure 2. Estimation of the measurement model

To test the mediating effect of detecting capability on the relationship between decentralization and organizational agility (H1) we test the two direct effects between decentralization and detecting capability as well as between detecting capability and organizational agility. Decentralization has a statistically significant, positive relationship with detecting capability (β = .541, p = .000), and detecting capability has a statistically significant, positive relationship with organizational agility (β = .679, p = .000). Subsequently, we test the direct effect between decentralization and organizational agility, and the indirect effect of decentralization on organizational agility via detecting capability. The direct relationship between decentralization and organizational agility is positive and significant (β = .300, p = .008). However, the inclusion of the mediator variable detecting capability leads to a statistically non-significant relationship between decentralization and organizational agility (β = -.139, p = .485). We conclude that the relationship between decentralization and organizational agility is fully mediated by detecting capability (confirming H1).

4.3 The Boundaries of IT Flexibility

Hypothesis 2 proposes a moderating effect of IT flexibility on the positive relationship between decentralization and detecting capability (H2a) and an indirect effect of decentralization on organizational agility through detecting capability conditional on IT flexibility (H2b). Thus, we report the p-value of the moderation and the mediation effect. Moreover, we follow the widely acknowledged argumentation, that since the sampling distribution of the product of two path coefficients is not normal, the sampling distribution of moderation and mediation effects is not normal either [33]. Therefore, using a t-test or a Sobel test would not be appropriate to evaluate the estimates' significance, because it requires a normal distribution. Consequently, we additionally use a non-parametric procedure to interpret the statistical significance of the moderation and mediation effects. The bias-corrected bootstrap confidence interval (CI) produces more accurate results than parametric tests (e.g., t-test) since it allows the interaction term and the indirect effect to be non-normally distributed. A particular effect is deemed statistically significant if the bias-corrected bootstrap confidence interval does not contain zero [33]. Based on these quality criteria, we find in our analysis:

First, the interaction effect between decentralization and IT flexibility on detecting capability is statistically significant, since the bootstrap confidence interval does not contain zero, and negative (H2a: $\beta = -.286$, 95% CI [-.648, -.022], $p = .057$). We find that IT flexibility acts as a moderator on the relationship between decentralization and detecting capability (confirming H2a).

Second, supporting the idea of a moderated mediation effect, the analysis is based on the index of moderated mediation [34]. We find the interaction effect between decentralization and IT flexibility on the mediated relationship (decentralization on organizational agility through detecting capability) to be also statistically significant and negative (H2b: $\beta = -.194$, 95% CI [-.454, -.009], $p = .075$). We find that IT flexibility acts as a moderator on the mediation effect of decentralization on organizational agility via detecting capability (confirming H2b).

Additionally, to provide a more nuanced analysis of H2b, we inspect the moderated mediation effect at various levels of IT flexibility. This procedure allows us to understand the nature of the moderated mediation effect in more detail [35]. For low levels of IT flexibility (one standard deviation below the mean) the effect of decentralization on organizational agility is high and significant (estimate = .534, 95% CI [.211, .968]). For levels of IT flexibility around the mean, the effect of decentralization on organizational agility through detecting capability is considerably lower and significant (estimate = .368, 95% CI [.150, .788]). Finally, for high levels of IT flexibility (one standard deviation above the mean) the effect of decentralization on organizational agility through detecting capability is considerably lower, but not significant (estimate = .201, 95% CI [-.024, .673]). To gain a deeper understanding of the magnitude and significance of the mediation effect at different levels of IT flexibility, we visualize the conditional indirect effect in Figure 3.

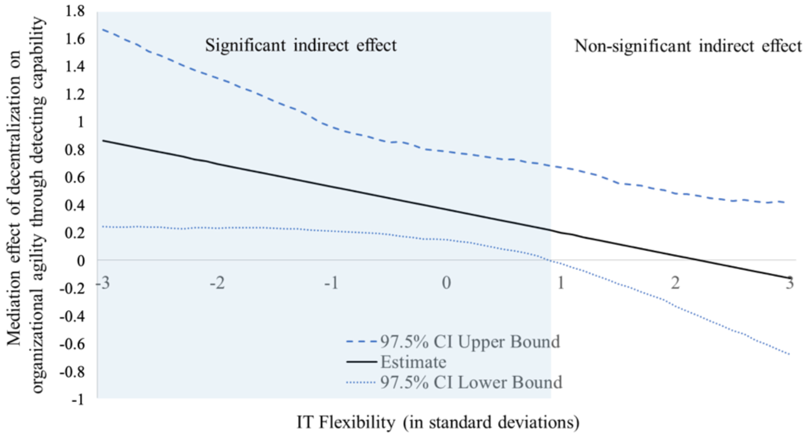


Figure 3. Mediation effect conditional on IT flexibility

A lower IT flexibility resulted in a higher mediation effect of decentralization on organizational agility through detecting capability. However, the mediation effect was only significant when IT flexibility was lower than +.8 standard deviation from the mean (shaded area in Figure 3). To safeguard the moderated mediation, we can report that the path coefficient between IT flexibility and detecting capability is positive and significant ($\beta = .361$, $p = .001$), and that the path coefficient between IT flexibility and organizational agility is statistically not significant ($\beta = .114$, $p = .325$).

5 Discussion

Our study supports earlier studies that show organizations with decentralized decision-making rights are better able to detect changes in their environment as the natural flow of innovative ideas is encouraged [18] and more employees are involved in the process of screening future changes [17]. Using this as a foundation, we provide two main contributions to the discussion from our analysis:

First, we uncover detecting capability as a mediator between decentralization and organizational agility. We show that the effect of decentralization on organizational agility is not only due to the fact that more employees can react to experienced changes by altering internal routines and processes independently. Instead, we find that the effect of decentralization on organizational agility is due to the increased capability to detect change. The decentralization of the organizational structure increases organizations' accumulated and shared knowledge about possible changes, which increases the organizations' ability to respond to these changes by appropriate actions. This provides a better understanding of how decentralization affects organizational agility.

Second, we explain the impact of IT flexibility on the relationship between decentralization and organizational agility via detecting capability. For organizations trying to improve their detecting capabilities by decentralizing decision-rights, high IT flexibility comes at a cost. Interestingly, too much flexibility of the IT function in combination with vertically distributed decision-rights decreases organizational agility. We attribute this to the fact that, if organizations have many employees at lower levels of the hierarchy allowed to make individual decisions, and at the same time an IT department flexible enough to satisfy the individual demands of those employees, organizations end up in a *complexity trap*. We find that decentralized decision-making combined with too many individual IT applications or components (provided through the flexible IT function) makes information and new solutions throughout the organizations less comparable. The *complexity trap* hinders the development of a shared understanding of future changes in the market and, hence, weakens organizations' agility. Nonetheless, we do not dismiss the widely acknowledged argumentation that IT capabilities have a positive effect on organizational agility [6], [11]. For example, Tiwana and Konsynski [24] found a positive interaction effect between IT governance decentralization and IT architecture modularity with regard to IT agility. Still, we provide theoretical and empirical evidence to the limits of IT flexibility. A highly flexible IT function in combination with highly decentralized decision rights decreases the capability to detect change, and, in turn, decreases also the ability to react to change. The *complexity trap* illustrates that providing many different IT applications in an organization with spread-out decision making, eliminates the potentially positive effects of the two measures due to complexity. Hence, we urge that potential harmful effects of (too extensive) IT flexibility and the complexity coming along with it should play a more prominent role in further organizational agility research. Additionally, further research should explore how decentralization and IT flexibility can be enhanced without increasing harmful complexity. We encourage research to analyze if a "fit" makes sense: the degree of decentralization must match the degree of IT flexibility.

As our study carries some substantial contributions, it also has some limitations that bear mentioning as well. First, our sample size is relatively small due to the fact that our study design required respondents to answer two surveys at different points in time. Second, our results indicate that a lower IT flexibility reduces the effect of decentralization on organizational agility through detecting capability. However, there is no statistically significant mediation effect when IT flexibility is very high (+.8 standard deviation from the mean). Strictly speaking, our results do not support the idea that a very high IT flexibility leads to a very low mediation effect. However, Figure 3 shows that the mediation effect conditional on the moderator is stable. Therefore, we are confident that our results are also valid for high levels of IT flexibility, and that the insignificance for very high levels of the moderator is rather due to the relatively small sample size and thereby a lack of statistical power. Third, our study relies only on data from the consulting service industry. Future research needs to extend our findings to different industries to validate our findings, to allow for generalization, and to find more explanations of the possible adverse effects IT flexibility.

6 Conclusion

In this study, we provide a theoretical explanation as well as empirical evidence for both how and why decentralization relates to organizational agility and to which degree this relationship depends on the IT flexibility. Based on our findings, we recommend managers undergoing a digital transformation to provide decentralized structures that empower employees to increase their understanding of possible changes that are likely going to affect the firm. This will increase organizations' ability to react to those changes, e.g., through improvements and additions to their internal processes.

However, we have to stress that decentralization in combination with a highly flexible IT should be treated with caution. We find that the effect of decentralization on the capability to detect changes is significantly reduced in organizations with a highly flexible IT function. We do not claim that IT flexibility is generally harmful to the detecting capability of organizations and thus agile organizations. However, managing the digital transformation of organizations requires to carefully consider the level of IT flexibility and decentralization to avoid the *complexity trap*. Furthermore, our analysis indicates that further research is necessary to investigate what causes possible drawbacks, and how IT flexibility and decentralization can be reconciled to support agility.

Appendix

Table 3. Overview Items and Loadings

Name	Item Label	Loading	Reference
DEC1	In your consultancy, managers and consultants can independently from top management decide about priority of projects and clients.	.718	[27]
DEC2	In your consultancy, managers and consultants can independently from top management decide about working overtime.	.594	
DEC3	In your consultancy, managers and consultants can independently from top management decide about employee recruitment and layoff	.732	
ITF1	Our IT unit/provider reacts flexibly to change requests from the consulting team.	.878	[15]
ITF2	The IT unit/provider realizes change requests from the consulting team in appropriate time.	.745	
ITF3	If there are critical bugs in our IT applications, they get fixed in a timely manner.	.667	
DET1	We seek novel approaches to satisfy future clients' needs.	.735	[28] [6]
DET2	The screening for unforeseeable changes is implemented in all our business activities.	.732	
DET3	We are constantly looking for opportunities to add new value to our clients.	.841	
OA1	We quickly capitalize changes and apparent chaos as new opportunities for services.	.686	[7]
OA2	We are quick to make and implement appropriate decisions in the face of client changes.	.715	
OA3	We are quick to react to new services launches by competitors with appropriate changes.	.804	

References

1. Yang, C., Liu, H.: Boosting Firm Performance via Enterprise Agility and Network Structure. Manag. Decis. 50, 1022–1044 (2012)
2. Westerman, G., Bonnet, D., McAfee, A.: Leading Digital: Turning Technology into Business Transformation. Harvard Business Review Press, Boston, MA (2014)
3. Tilson, D., Lyytinen, K., Sørensen, C.: Digital Infrastructures: The Missing IS Research Agenda. Inf. Syst. Res. 21, 1–12 (2010)
4. Worley, C.G., Lawler, E.E.: Agility and Organization Design: A Diagnostic Framework. Organ. Dyn. 39, 194–204 (2010)
5. Burns, T., Stalker, G.M.: The Management of Innovation. Tavistock, London (1961)
6. Sambamurthy, V., Bharadwaj, A., Grover, V.: Shaping Agility through Digital Options: Reconceptualizing the Role of IT in Contemporary Firms. MIS Q. 27, 237 (2003)
7. Lu, Y., Ramamurthy, K.: Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination. MIS Q. 35, 931 (2011)
8. Byrd, T.A., Turner, D.E.: An Exploratory Examination of the Relationship between Flexible IT Infrastructure and Competitive Advantage. Inf. Manag. 39, 41–52 (2001)
9. Teece, D., Peteraf, M.A., Leih, S.: Dynamic Capabilities and Organizational Agility. Calif. Manage. Rev. 58, 13–35 (2016)
10. Bernardes, E.S., Hanna, M.D.: A Theoretical Review of Flexibility, Agility and Responsiveness in the Operations Management Literature. Int. J. Oper. Prod. Manag. 29, 30–53 (2009)

11. Overby, E., Bharadwaj, A., Sambamurthy, V.: Enterprise Agility and the Enabling Role of Information Technology. Eur. J. Inf. Syst. 15, 120–131 (2006)
12. Zhang, Z., Sharifi, H.: A Methodology for Achieving Agility in Manufacturing Organisations. Int. J. Oper. Prod. Manag. 20, 496–513 (2000)
13. Tallon, P.P., Pinsonneault, A.: Competing Perspectives on the Link Between Strategic IT Alignment and Org. Agility: Insights from a Mediation Model. MIS Q. 35, (2011)
14. Mintzberg, H.: Structure in 5's: A Synthesis of the Research on Organization Design. Manage. Sci. 26, 322–341 (1980)
15. Martin, S.F., Wagner, H.-T., Beimborn, D.: Process Documentation, Operational Alignment, and Flexibility in IT Outsourcing Relationships: A Knowledge-Based Perspective. In: ICIS. pp. 1–12 (2008)
16. Winkler, T.J., Wessel, M.: A Primer on Decision Rights in Information Systems : Review and Recommendations. In: ICIS. pp. 1–17 (2018)
17. Dewar, R., Werbel, J.: Universalistic and Contingency Predictions of Employee Satisfaction and Conflict. Adm. Sci. Q. 24, 426 (1979)
18. Hage, J., Dewar, R.: Elite Values Versus Organizational Structure in Predicting Innovation. Adm. Sci. Q. 18, 279 (1973)
19. Pierce, J.L., Delbecq, A.L.: Organization Structure, Individual Attitudes and Innovation. Acad. Manag. Rev. 2, 27–37 (1977)
20. Van Assen, M.F.: Agile-based Competence Management: The Relation between Agile Manufacturing and Time-based Competence Management. Int. J. Agil. Manag. Syst. 2, 142–155 (2000)
21. Bennett, C.A.: The Human Factors of Work. J. Hum. Factors Ergon. Soc. 15, 281–287 (1973)
22. Miller, D., Friesen, P.H.: Strategy-Making and Environment: The Third Link. Strateg. Manag. J. 4, 221–235 (1983)
23. Mocker, M.: What Is Complex About 273 Applications? Untangling Application Architecture Complexity in a Case of European Investment Banking. In: HICSS. pp. 1–14 (2009)
24. Tiwana, A., Konsynski, B.: Complementarities Between Organizational IT Architecture and Governance Structure. Inf. Syst. Res. 21, 288–304 (2010)
25. Nissen, V.: Digital Transformation of the Consulting Industry. Springer Publishing (2018)
26. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. J. Appl. Psychol. 88, 879–903 (2003)
27. Claver-Cortés, E., Pertusa-Ortega, E.M., Molina-Azorín, J.F.: Characteristics of Organizational Structure relating to Hybrid Competitive Strategy. J. Bus. Res. 65, (2012)
28. Ahsan, M., Ngo-ye, L.: The Relationship Between IT Infrastructure and Strategic Agility in Organizations. In: AMCIS. pp. 415–427 (2005)
29. Armstrong, J.S., Overton, T.S.: Estimating Nonresponse Bias in Mail Surveys. J. Mark. Res. 14, 396 (1977)
30. Albright, J.J., Park, H.M.: Confirmatory Factor Analysis using Amos, LISREL, Mplus, SAS/STAT CALIS. Work. Pap. Indiana Univ. (2009)
31. Hu, L., Bentler, P.M.: Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Struct. Equ. Model. 6, 1–55 (1999)
32. Cheung, G.W., Lau, R.S.: Accuracy of Parameter Estimates and Confidence Intervals in Moderated Mediation Models. Organ. Res. Methods. 20, 746–769 (2017)
33. Williams, J., MacKinnon, D.P.: Resampling and Distribution of the Product Methods for Testing Indirect Effects in Complex Models. Struct. Equ. Model. 15, 23–51 (2008)
34. Hayes, A.F.: An Index and Test of Linear Moderated Mediation. Multivariate Behav. Res. 50, 1–22 (2015)
35. Preacher, K.J., Rucker, D.D., Hayes, A.F.: Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. Multivariate Behav. Res. 42, 185–227 (2007)

Platform Openness: A Systematic Literature Review and Avenues for Future Research

David Soto Setzke[1], Markus Böhm[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Munich, Germany
{setzke,markus.boehm,krcmar}@in.tum.de

Abstract. Open platforms such as Facebook or Android have stimulated innovation and competition across industries. Information systems literature has analyzed platforms from a variety of perspectives. The aim of this paper is to synthesize and integrate extant interdisciplinary research on the concept of platform openness. Towards this end, we conducted a literature review and analyzed the results with deductive and inductive coding approaches. We identified five distinct themes: measurement frameworks, implementation mechanisms, drivers for opening and closing platforms, trade-offs in designing openness, and the impact of changing openness on ecosystems. We propose three avenues for future research: finding the optimal degree of platform openness, integrating perspectives on accessibility and transparency, and analyzing the influence of openness and other factors with configurational theories. This paper contributes to research on platforms by laying out the main themes and perspectives in the research stream of platform openness and by identifying areas for future research.

Keywords: Platform Openness, Digital Platforms, Platform Ecosystems

1 Introduction

Digital platforms have transformed entire industries by leveraging the concept of open innovation [1] and have stimulated generativity[1] and competition [2, 3]. The cases of the social network platforms MySpace and Facebook are prototypical examples for the competitive advantages of open platform strategies. While MySpace kept their system closed, trying "to create every feature in the world" [4] on their own, Facebook decided in 2007 to open themselves to a worldwide pool of third-party developers, allowing them to build applications on top of the social networking platform [5]. Six months later, 8.000 third-party applications had been added and one year later, Facebook surpassed MySpace in terms of unique monthly visitors [6, 7]. When Apple initially released the iPhone with its iOS[2] operating system in 2007, it was closed to external developers but soon after, Apple released an official Software Development Kit (SDK)

[1] We refer to generativity as "a technology's overall capacity to produce unprompted change driven by large, varied, and uncoordinated audiences" [2].
[2] Until 2010: iPhone OS

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

and set up a distribution channel for third-party applications, the Apple AppStore [8]. Google's Android operating system entered the market of mobile platforms later but was released under an open source license and came with a less restrictive application marketplace. [9]. In 2010, Android first surpassed iOS in terms of worldwide smartphone sales and has remained the dominating mobile platform since then (with a market share of 87.8 % as of 2017) [10].

These real-world examples show the strategic role played by platform openness. Information Systems (IS) literature has analyzed the phenomenon of platforms from a variety of perspectives [3]. The concept of platform openness is commonly referred to as placing restrictions on the development, commercialization, or use of a platform [11]. More specifically, platform openness is controlled by platform owners through the use of platform governance mechanisms, such as "deliberate regulations and rules about access and boundary control" [12]. In IS literature, platform openness has started to gain traction in the last years but each study focuses on different aspects [11-17]. This is aggravated by the fact that relevant insights are also to be found in the neighboring literature streams of management [18-22] and computer science [23-25]. Hence, IS research lacks an integrated view of different, inter-disciplinary perspectives on platform openness. Due to the fragmentation of knowledge on platforms, scholars have called for consolidating extant research perspectives (see, e.g., de Reuver, Sørensen and Basole [3]).

The purpose of this paper is to synthesize the current state of research and to integrate different perspectives on platform openness in IS literature and neighboring literature streams. To this end, we conducted a systematic literature review and analyzed the resulting publications with deductive and inductive coding approaches. We find that literature focuses on technological accessibility but neglects the perspective of transparency. Furthermore, we identify five distinct themes: measuring platform openness, mechanisms for implementing openness, drivers for opening up or closing down, trade-offs in designing the degree of openness, and the impact of changing degrees of openness on platform-centric ecosystems.

The remainder of this paper is structured as follows. The first section describes the design of the literature review and the employed coding approach. The second section structures the analyzed publications with a deductive coding scheme based on different research perspectives. Subsequently, we present and discuss the identified research themes. Finally, we present and discuss areas for future research and conclude the review.

2 Design of the literature review

For conducting our literature review, we followed the guidelines of Webster and Watson [26]. Drawing on the typology of literature reviews developed by Paré, Trudel, Jaana and Kitsiou [27], our review constitutes a descriptive review since our goal was to synthesize and represent the current state of the art of research on platform openness. We restricted our review to the openness of digital platforms following the conceptualization of de Reuver, Sørensen and Basole [3] as "purely technical artefacts where the platform is an extensible codebase, and the ecosystem comprises third-party modules complementing this codebase". We focused on the journals included in the *AIS Senior Scholars' Basket of Journals*. To include the perspective of management, we also selected the journals *Management Science* (MS) and *Organization Science* (OS). We conducted a search with the term "platform AND open*" on titles, abstracts, and keywords and screened the abstract of 53 publications, resulting in eleven selected articles. If the relevance for our review was unclear after reading the abstract, we read the full article. In a second step, we extended our search to highly ranked IS conferences and the IEEE Explore Digital Library to include the perspective of computer science. We restricted our search to the more specific term "platform AND openness" in order to get a manageable set of publications, resulting in 685 potentially relevant articles. Again, we screened the abstracts in order to decide whether to include the article, resulting in 14 selected publications. Afterwards, we performed a forward and backward search on the articles that were selected so far, leading to the inclusion of another 48 articles. Finally, our sample comprised 73 relevant articles (see **Table 1**).

Table 1. Summary of the literature search process

Outlet		*Search*	*Hits*	*Selected*
Top journals	AIS Basket of Eight	*"platform AND open*"* in Title | Abstract | Keywords	34	8
	OS		8	1
	MS		11	2
IS conferences	ICIS	*"platform AND openness"* in Title | Abstract | Keywords	28	5
	ECIS		24	3
	PACIS		12	3
	HICSS		351	2
	WI		7	0
	AMCIS		35	0
Other outlets	IEEE Xplore		228	1
	Journals	Forward and backward search	-	20
	Conferences		-	18
	Dissertations		-	2
	Books		-	4
	Other		-	4
Total			738	73

In a next step, we iteratively coded the articles, using both a deductive and inductive approach [28]. Our deductive scheme was adapted from the guidelines of Bandara, Furtmueller, Gorbacheva, Miskon and Beekhuyzen [29] and comprised definition and measurement frameworks of platform openness, employed research methodologies, future work, and distinct levels of openness. Regarding our inductive approach, we engaged in open coding, axial coding, and selective coding to capture and distill concepts emerging from our sample of publications [28]. Based on 50 definitions of platform openness from our article sample, we clustered recurrent themes and extracted three distinct levels and two dimensions in order to classify extant research perspectives. Furthermore, five distinct research themes emerged throughout our open coding process[3]. Based on our classification and the identified themes, we derived promising avenues for future research. **Table 2** gives an overview of the results of our literature review, our approach to generate these results, and the respective section of this paper.

Table 2. Overview of the results of this literature review

Section	*Results*	*Approach*
3	Classification of extant research	Inductive and deductive coding of identified literature
4	Central themes in extant research	Inductive and deductive coding of identified literature
5	Avenues for future research	Analysis of classification table from section 3 and identification of unanswered questions in central themes from section 4

3 Research perspectives on platform openness

Of our analyzed articles, 68 % employ an explicit definition of platform openness (see, e.g., Eisenmann, Parker and Van Alstyne [11], Boudreau [18], Anvaari and Jansen [23], Arakji and Lang [31]) while only 30 % use or introduce a qualitative or quantitative framework for measuring openness (see, for example, Benlian, Hilkert and Hess [13], Ondrus, Gannamaneni and Lyytinen [17], Anvaari and Jansen [23]). In terms of research methodologies, we distinguish between qualitative, quantitative, mixed, design science, and conceptual research approaches [29]. A 33 % of the papers are based on qualitative research methods, such as single or multiple case studies conducted with, for example, app stores [14, 32] or mobile payment platforms [33, 34]. Quantitative research methods are employed in 31 % of the papers, comprising mostly econometric analyses [7, 15], surveys [35], and simulations [31, 36]. Conceptual and mixed approaches are represented with 19 % [37] and 13 % [38], while design science strategies are only used in 4 % of the analyzed publications [16].

Based on different definitions of platform openness, we identified three distinct levels and two dimensions. Openness can be implemented on three levels: organization, technology, and users. The organizational level "relates to the strategic involvement of key stakeholders who control the platform and provide the platform services to different user groups" [17]. The technology level refers to the provisioning of "technical means

[3] For instance, the theme "trade-offs in designing the degree of openness" was derived from codes such as "Decisions to open a platform entail tradeoffs between adoption and appropriability" [11] or "[..] it may be a trade-off between attracting a developer community […] and ensuring high standards" [30].

for complementary providers (i.e. companies that provide alternative technology, products or services for the platform) to access the core functions of the platform" [39].

Table 3. Identified research perspectives on platform openness

Article	*Organization*		*Technology*		*Users*	
	Acces-sibility	Trans-parency	Acces-sibility	Trans-parency	Acces-sibility	Trans-parency
Top journals and IS conferences						
Benlian, Hilkert and Hess [13]			X	X	X	X
Boudreau [18]	X		X			
Boudreau [20]			X			
Casadesus-Masanell and Llanes [7]			X		X	
Foerderer, Schuetz and Kude [40]			X	X		
Furstenau and Auschra [41]			X		X	
Gawer [37]			X			
Ghazawneh and Henfridsson [14]					X	X
Hilkert, Benlian, Sarstedt and Hess [42]			X	X	X	X
Karhu, Gustafsson and Lyytinen [43]			X			
Kazan and Damsgaard [34]	X				X	
Kuebel and Zarnekow [44]	X		X		X	X
Kwon, Oh and Kim [36]					X	
Niculescu, Wu and Xu [45]			X	X		
Nikou, Bouwman and de Reuver [46]			X			
Ondrus, Gannamaneni and Lyytinen [17]	X		X		X	
Park, Lee and Lee [47]			X		X	
Parker and Van Alstyne [48]			X		X	
Parker and Van Alstyne [21]			X		X	
Parker, Van Alstyne and Jiang [15]			X		X	
Schreieck, Wiesche and Krcmar [49]			X			
Song, Baker, Wang, Choi and Bhattacherjee [38]			X			
Wessel, Thies and Benlian [12]					X	
West [50]	X	X	X		X	
Other papers						
	14	6	54	11	33	5
Total articles	19	7	62	15	40	9

On the user level, openness "is defined by the level of discrimination that the platform exercises against different segments of the potential customer base" [17]. Furthermore, openness can be categorized as either providing accessibility or transparency [23, 44, 51]. Accessibility focuses on the degree of discrimination against different roles and determines whether providers, third-party developers, or end users are allowed to join and access the platform [23, 51, 52]. Transparency, on the other hand, relates to the "understanding of what is happening and why" and thus determines whether platform-related governance decisions are comprehensible [23, 51]. On each of these levels, a platform can be open or closed. Furthermore, for each of these levels, the platforms' degrees of accessibility and transparency can be determined. On the technology level, for example, accessibility refers to the degree to which third-party developers are allowed to contribute to the platform by building new applications. Transparency, on the other hand, refers to the degree to which it is made understandable to these developers how and under what conditions third-party applications can be created and distributed through channels like the platform's application marketplace. Similarly, on the user level, accessibility reflects the possibility for users to participate on a platform (such as Uber), while transparency refers to how and to what extent the rules for participating are made comprehensible. The resulting coding matrix of our publication sample shows that most papers focus on accessibility on the technology and user level, while the dimension of transparency on all the levels is mostly neglected, especially on the user and organization level (see **Table 3**).

4 Central themes in research on platform openness

4.1 Measuring platform openness

Platform openness should not be measured as a binary variable, but rather depicted as a continuum [50]. As already introduced earlier, Eisenmann, Parker and Van Alstyne [11] distinguish between four distinct roles in ecosystems (sponsors, providers, complementors, and end users) towards which platforms can be open or closed. Still, even in platforms that are seemingly open towards a specific role, openness may still be restricted to a certain degree. The source code of the operating system Linux, for example, is accessible to everyone, but contributors need to adhere to strict governance processes comprising code review and quality appraisal [18, 24]. The framework of architectural openness developed by Anvaari and Jansen [23] considers this distinction. The architecture of a platform is divided into four layers: kernel, middleware, native applications, and extended applications. The framework shows whether it is possible to modify, extend, or integrate each layer and whether permission by the platform owner is needed for these activities.

Other frameworks focus on specific architectural aspects. Schlagwein, Schoder and Fischbach [16] propose a matrix-based framework for measuring the openness of platform resources along the dimensions of access and control. Access to resources can be exclusive, on a group-basis, or open. Control of resources can be exercised by the platform owner, by a group, or by an external actor. Ghazawneh and Henfridsson [14] focus on distribution channels and present a typology for digital application marketplaces. They distinguish between closed, censored, focused, and open marketplaces with different regulatory designs. Taking the perspective of complementors, Benlian, Hilkert and Hess [13] develop an instrument for measuring complementors' perceived platform openness along the dimensions of technical platform, distribution channel, accessibility, and transparency.

4.2 Mechanisms for implementing openness

For structuring mechanisms for implementing openness, we draw on the notion of horizontal and vertical openness by Eisenmann, Parker and Van Alstyne [11]. Horizontal openness refers to allowing rival platform's users to interact with the own platform or allowing additional parties to participate in the platform's commercialization or technical development. Vertical openness refers to granting third-party developers access to resources for developing complementary applications.

For implementing horizontal openness, platform owners choose to establish interoperability with other platforms in order to increase their market potential, either as part of a competitive or collaborative strategy (see, for example, the interoperability agreement between the instant messaging services of Yahoo and Microsoft) [11, 17]. Another strategy, that is especially attractive for mature platforms, consists of licensing the own platform to additional providers while retaining control over the platform's technology (see, e.g. Microsoft Windows) [11]. Going even further, platform sponsors may also give up ownership over technology and invite partners for joint sponsorship and development (see, for example, the Linux operating system or other open source software projects) [11, 18, 50].

Vertical openness is implemented through boundary resources [49, 53], i.e. the "software tools and regulations that serve as the interface for the arm's length relationship between the platform owner and the application developer" [8]. In practice, this includes technical boundary resources such as Application Programming Interfaces (APIs), SDKs and non-technical boundary resources such as technical documentation and support or the provided community [54-56]. Furthermore, distribution channels such as app stores are offered to facilitate the diffusion of third-party complements [14]. From a policy perspective, platform owners can restrict access to resources, e.g., by charging usage fees or by reserving access to selected groups of developers [18, 57]. In addition, they can exercise content control on distribution channels through prescreening, review, and approval processes [14, 58, 59].

4.3 Drivers for opening up or closing down

Platform owners decide to open up platform boundaries when seeking to stimulate growth by increasing their user base [17]. A larger end user base leads to higher market shares while a higher developer base allows the platform owner to access external resources and stimulate innovation even or especially when lacking own competencies to innovate [60, 61]. In the case where changing the level of openness is complicated through technological or cultural constraints owners tend to more liberally open the platform when expecting an increasing developer base [15, 17]. The need to comply with or the uncertainty about legal regulations may also be factors to open or close a platform (see, e.g., the lawsuit provoked by Microsoft's decision to bundle Windows with Internet Explorer) [41, 53, 60].

The degree of openness is not a fixed, static choice, but may vary over time, shifting from closed to open or vice-versa [30, 62]. Platform-to-platform competition, for example, where each platform intends to attract more developers may incentivize platform owners to increase openness [7, 22]. On the other hand, certain features of a platform may become so valuable over time that the platform owner does not gain any more benefits by keeping these parts fully open [15]. This can be observed at the practical examples of platforms such as LinkedIn, Twitter, or Instagram. In 2015, all three of these platforms announced the discontinuation of a large portion of their formerly open APIs, mentioning, among others, competitive threads to their businesses [63-65]. This led to the shutdown of several third-party application who could not afford the transition to the companies' partner programs or whose application use cases did not meet new terms of service [66].

4.4 Trade-offs in designing the degree of openness

Two central trade-offs need to be balanced by platform owners: adoption vs. appropriability[4] [15, 50] and diversity vs. control [18, 30, 68]. First, as already shown, higher openness leads to adoption by complementary developers. Higher openness however also reduces switching costs and increases inter-platform competition, thus making it more difficult to appropriate profits [50]. Second, higher openness leads to more diversity of complementary applications through open innovation. On the other hand, the platform owner may lose control over the quality of applications and be faced with complex coordination of resources and strategic interests [18, 20]. During the so-called "Atari shock", for example, a high number of low-quality games for the video gaming platform Atari that exercised no content control at all led to its eventual demise [69].

4.5 Impact of changing degrees of openness on platform-centric ecosystems

On the sponsor level, higher platform openness leads to the necessity of increasing modularity and more complex system architectures on the technology level [70]. In collectively sponsored platforms, increased openness on the sponsor layer may be a source of conflict resulting from deciding on the inclusion of new sponsors [33]. Sponsors and providers may also benefit from lower development costs through effectively outsourcing innovation [15, 71].

Several qualitative and quantitative studies show that higher openness leads to increasing adoption among complementors and a high quantity and variety of complementary applications [7, 17, 19, 44, 59, 72-75]. Puvvala, Dutta, Roy and Seetharaman [54] support these results and show the importance of provisioning tooling and reasonable license costs. A case study on the crowdfunding platform Kickstarter

[4] We refer to appropriability as "the ability of different stakeholders to retain for themselves the financial benefits that arise through the exploitation of an innovation" [67].

shows that increased openness on the complementor side may lead to a destabilized ecosystem [12]. After relaxing the screening processes for new campaigns on their platform, campaign success rates decreased and competition between project creators increased because of an altered ratio of campaigns to backers. A particular challenge lies in determining the right degree of openness. While granting access to complementary developers is associated with a rising innovation rate, after a certain threshold the rate decreases again in a curvilinear manner due to excessive competition between developers [18, 20].

On the user level, Müller, Kijl and Martens [32] argue that stricter content control leads to higher quality of third-party applications but on the other hand, higher competition induced by low control also leads to lower prices for end users. In terms of end user adoption, Hagiu [76] and Moon and Choi [9] suggest that lower openness may induce higher use adoption due to increasing competition. Finally, the openness towards third-party developers does not influence adoption among end consumers, as shown by Nikou, Bouwman and de Reuver [46].

5 Avenues for future research on platform openness

In this section, we point out and discuss central avenues for future research that appear promising based on our literature review. First, we call for further research on finding the optimal degree of platform openness. Second, we suggest integrating perspectives on accessibility and transparency. Third, we discuss the adoption of novel research methodologies in the context of organizational and technical configurations and the role of platform openness.

The findings of Boudreau [18] and Parker, Van Alstyne and Jiang [15] characterize the relationship between innovation and openness as curvilinear, suggesting that platform openness can be optimized [13]. However, little is known about the factors that influence the threshold at which innovation decreases again. The evidence presented by Boudreau [18] is based only on data on handheld computing systems from 1990-2004 and has since then not been verified nor replicated using data on more recent platforms. As of today, recent examples of platforms with varying degrees of openness (see, as already discussed: Hofer-Shall [63], Instagram [64], Trachtenberg [65]) provide data that allow for reexamining the question of optimal openness and its accompanying conditions. The results could be valuable for theoretical advances on platform research as well as for practical guidelines on effective platform governance.

As our coding has shown, few articles consider the transparency dimension on openness, such as technical documentation, communication with end users, or transparency of market mechanisms. Yet research has demonstrated that aspects of transparency considerably influence platform adoption among complementors [54, 77, 78]. Hence, integrating perspectives on accessibility and transparency regarding platform openness promises to be a fruitful research area. For example, different best practices regarding the implementation and promotion of transparency could be identified through a multiple-case analysis of successful platforms. This could yield insights on the design of successful platform ecosystems for end users and complementors, ultimately resulting in higher platform adoption.

Several studies have identified and discussed drivers and impacts of changing degrees of openness. However, we argue that the complex causal interplay of these drivers and organizational and technical preconditions in the firms and platforms influence the degree of openness, rather than stern linear relationships (see Vis [79] for a detailed discussion). For this reason, we call for the use of research methods that take into account equifinality and complex non-linear relationships, such as fuzzy-set Qualitative Comparative Analysis (fsQCA) [80]. FsQCA with platforms as unit of analysis has been employed by, for example, Dellermann and Reck [81], Dellermann and Reck [82] and Dellermann, Jud and Reck [83] for analyzing user loyalty, platform governance, and perceived risk. Future research could examine the effect of the interplay of openness and other factors such as the number of sides or the amount of partners on successful or unsuccessful platform launches [84], deriving relevant insights for practitioners.

6 Conclusion

The goal of this paper was to synthesize the current state of research on platform openness and to identify avenues for future research. Literature analyzes platform openness on different levels and dimensions, but neglects aspects of transparency. The main themes comprise measurement frameworks, implementation mechanisms, drivers for opening and closing platforms, trade-offs in designing openness, and the impact of changing openness on ecosystems. Based on our results, we propose three distinct issues for future research: finding the optimal degree of platform openness, integrating perspectives on accessibility and transparency, and analyzing the interplay of openness and other factors with novel research methods.

References

1. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting From Technology. Harvard Business Press (2006)
2. Zittrain, J.L.: The Generative Internet. Harvard Law Review 1974-2040 (2006)
3. de Reuver, M., Sørensen, C., Basole, R.C.: The Digital Platform: A Research Agenda. Journal of Information Technology (2017)
4. Gillette, F.: The Rise and Inglorious Fall of MySpace. Bloomberg Businessweek, vol. June, pp. 52-59 (2011)
5. Jung, G., Lee, B.: How Did Facebook Outspace Myspace With Open Innovation? An Analysis Of Network Competition With Changes Of Network Topology. 15th Pacific Asia Conference on Information Systems, Brisbane, Australia (2011)
6. Torkjazi, M., Rejaie, R., Willinger, W.: Hot Today, Gone Tomorrow: On the Migration of MySpace Users. 2nd ACM workshop on Online social networks, pp. 43-48. ACM, Austin, TX, USA (2009)
7. Casadesus-Masanell, R., Llanes, G.: Investment Incentives in Open-Source and Proprietary Two-Sided Platforms. Journal of Economics & Management Strategy 24, 306-324 (2015)

8. Ghazawneh, A., Henfridsson, O.: Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model. Information Systems Journal 23, 173-192 (2013)
9. Moon, J., Choi, S.: Impact of Mobile Platform Strategy on Platform Generativity and Competition. Academy of Strategic Management Journal 15, 47-61 (2016)
10. Statista: Marktanteile der führenden Betriebssysteme am Absatz von Smartphones weltweit vom 1. Quartal 2009 bis zum 2. Quartal 2017. https://de.statista.com/statistik/daten/studie/73662/umfrage/marktanteil-der-smartphone-betriebssysteme-nach-quartalen/ (Accessed: 25/03/2018)
11. Eisenmann, T., Parker, G., Van Alstyne, M.: Opening Platforms: How, When and Why? In: Gawer, A. (ed.) Platforms, Markets and Innovation. Edward Elgar Publishing Limited, Cheltenham, UK (2009)
12. Wessel, M., Thies, F., Benlian, A.: Opening the Floodgates: The Implications of Increasing Platform Openness in Crowdfunding. Journal of Information Technology 32, 344-360 (2017)
13. Benlian, A., Hilkert, D., Hess, T.: How Open is This Platform? The Meaning and Measurement of Platform Openness from the Complementors' Perspective. Journal of Information Technology 30, 209-228 (2015)
14. Ghazawneh, A., Henfridsson, O.: A Paradigmatic Analysis of Digital Application Marketplaces. Journal of Information Technology 30, 198-208 (2015)
15. Parker, G., Van Alstyne, M., Jiang, X.: Platform Ecosystems: How Developers Invert the Firm. MIS Quarterly 41, 255-266 (2017)
16. Schlagwein, D., Schoder, D., Fischbach, K.: Openness of Information Resources – A Framework-based Comparison of Mobile Platforms. 18th European Conference on Information Systems, Pretoria, South Africa (2010)
17. Ondrus, J., Gannamaneni, A., Lyytinen, K.: The Impact of Openness on the Market Potential of Multi-Sided Platforms: A Case Study of Mobile Payment Platforms. Journal of Information Technology 30, 260-275 (2015)
18. Boudreau, K.: Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. Management Science 56, 1849-1872 (2010)
19. Boudreau, K.: Platform Boundary Choices & Governance: Opening-Up While Still Coordinating and Orchestrating. In: Furman, J., Gawer, A., Silverman, B.S., Stern, S. (eds.) Advances in Strategic Management: Entrepreneurship, Innovation, and Platforms, vol. 37. Emerald Publishing Limited, 227 - 297 (2017)
20. Boudreau, K.J.: Let a Thousand Flowers Bloom? An Early Look at Large Numbers of Software App Developers and Patterns of Innovation. Organization Science 23, 1409-1427 (2012)
21. Parker, G., Van Alstyne, M.: Innovation, Openness, and Platform Control. Management Science (2017)
22. Parker, G., Van Alstyne, M.: Platform Strategy. In: Augier, M., Teece, D. (eds.) Palgrave Encyclopedia of Strategic Management. Palgrave MacMillan, London, UK (2014)
23. Anvaari, M., Jansen, S.: Evaluating Architectural Openness in Mobile Software Platforms. Fourth European Conference on Software Architecture: Companion Volume, Copenhagen, Denmark (2010)
24. Berger, T., Pfeiffer, R.-H., Tartler, R., Dienst, S., Czarnecki, K., Wąsowski, A., She, S.: Variability Mechanisms in Software Ecosystems. Information and Software Technology 56, 1520-1535 (2014)
25. Teixeira, J.: On the Openness of Digital Platforms/Ecosystems. 11th International Symposium on Open Collaboration, San Francisco, CA, USA (2015)

26. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, xiii-xxiii (2002)
27. Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: A typology of literature reviews. Information & Management 52, 183-199 (2015)
28. Wolfswinkel, J.F., Furtmueller, E., Wilderom, C.P.M.: Using Gounded Theory as a Method for Rigorously Reviewing Literature. European Journal of Information Systems 22, 45-55 (2017)
29. Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., Beekhuyzen, J.: Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support. Communications of the Association for Information Systems 37, 154-204 (2015)
30. Tåg, J.: Open Versus Closed Platforms. IFN Working Paper. Research Institute of Industrial Economics Stockholm, Sweden (2008)
31. Arakji, R., Lang, K.: An Evolutionary Theory of Innovation and Strategic Platform Openness for Web 2.0 Businesses. Proceedings of JAIS Theory Development Workshop, Sprouts: Working Papers on Information Systems, vol. 10, pp. 10-85 (2010)
32. Müller, R.M., Kijl, B., Martens, J.K.J.: A Comparison of Inter-Organizational Business Models of Mobile App Stores: There is more than Open vs. Closed. Journal of Theoretical and Applied Electronic Commerce Research 6, 13-14 (2011)
33. de Reuver, M., Verschuur, E., Nikayin, F., Cerpa, N., Bouwman, H.: Collective Action For Mobile Payment Platforms: A Case Study on Collaboration Issues Between Banks and Telecom Operators. Electronic Commerce Research and Applications 14, 331-344 (2015)
34. Kazan, E., Damsgaard, J.: Towards A New Market Entry Framework For Digital Payment Platforms. Communications of the Association for Information Systems 38, 761-783 (2016)
35. Schaarschmidt, M., Kilian, T.: Peripheral Motivation and Creativity in Controlled Platforms: An Analysis Based on Facebook and iPhone Application Developers. 20th European Conference on Information Systems, Barcelona, Spain (2012)
36. Kwon, H.E., Oh, W., Kim, T.: Platform Structures, Homing Preferences, and Homophilous Propensities in Online Social Networks. Journal of Management Information Systems 34, 768-802 (2017)
37. Gawer, A.: Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework. Research Policy 43, 1239-1249 (2014)
38. Song, J., Baker, J., Wang, Y., Choi, H.Y., Bhattacherjee, A.: Platform Adoption By Mobile Application Developers: A Multimethodological Approach. Decision Support Systems (2018)
39. Nikayin, F., De Reuver, M., Itala, T.: Collective Action for a Common Service Platform for Independent Living Services. International Journal of Medical Informatics 82, 922-939 (2013)
40. Foerderer, J., Schuetz, S.W., Kude, T.: Add-on Solution Success: A Configurational View on Knowledge Sharing in Digital Platforms. 35th International Conference on Information Systems, Auckland, CA, USA (2014)
41. Furstenau, D., Auschra, C.: Open Digital Platforms in Health Care. 37th International Conference on Information Systems, Dublin, Ireland (2016)
42. Hilkert, D., Benlian, A., Sarstedt, M., Hess, T.: Perceived Software Platform Openness: The Scale and its Impact on Developer Satisfaction. 32nd International Conference on Information Systems, Shanghai, China (2011)
43. Karhu, K., Gustafsson, R., Lyytinen, K.: Exploiting and Defending Open Digital Platforms with Boundary Resources: Android's Five Platform Forks. Information Systems Research 29, 479-497 (2018)

44. Kuebel, H., Zarnekow, R.: Exploring Platform Adoption in the Smart Home Case. 36th International Conference on Information Systems, Fort Worth, Texas, USA (2015)
45. Niculescu, M.F., Wu, D.J., Xu, L.: Strategic Intellectual Property Sharing: Competition on an Open Technology Platform Under Network Effects. Information Systems Research 29, 498-519 (2018)
46. Nikou, S., Bouwman, H., de Reuver, M.: A Consumer Perspective on Mobile Service Platforms: A Conjoint Analysis Approach. Communications of the Association for Information Systems 34, 1409-1424 (2014)
47. Park, Y., Lee, G., Lee, Z.: EA Configurations: Interplay of EA Design Factors, Strategy Types, and Environments. 34th International Conference on Information Systems, Milan, Italy (2013)
48. Parker, G., Van Alstyne, M.: Managing Platform Ecosystems. 29th International Conference on Information Systems, Paris, France (2008)
49. Schreieck, M., Wiesche, M., Krcmar, H.: The Platform Owner's Challenge to Capture Value – Insights from a Business-to-Business IT Platform. 38th International Conference on Information Systems, Seoul, South Korea (2017)
50. West, J.: How Open is Open Enough? Research Policy 32, 1259-1285 (2003)
51. West, J., O'Mahony, S.: The Role of Participation Architecture in Growing Sponsored Open Source Communities. Industry and Innovation 15, 145-168 (2009)
52. Kuebel, H., Hanner, N., Zarnekow, R.: An Expert View on the Role of Complementary Assets for the Adoption of Smart Home Platforms. 19th Pacific Asia Conference on Information Systems, Singapore (2015)
53. Schreieck, M., Wiesche, M.: How Established Companies Leverage IT Platforms For Value Creation - Insights From Banking. 25th European Conference on Information Systems, Guimarães, Portugal (2017)
54. Puvvala, A., Dutta, A., Roy, R., Seetharaman, P.: Mobile Application Developers' Platform Choice Model. 49th Hawaii International Conference on System Sciences, Koloa, HI, USA (2016)
55. Dal Bianco, V., Myllarniemi, V., Komssi, M., Raatikainen, M.: The Role of Platform Boundary Resources in Software Ecosystems: A Case Study. IEEE/IFIP Conference on Software Architecture (WICSA). IEEE, Sydney, Australia (2014)
56. Tiwana, A.: Platform Ecosystems: Aligning Architecture, Governance, and Strategy. Morgan Kaufmann, Waltham, MA, USA (2014)
57. Homscheid, D., Kilian, T., Schaarschmidt, M.: Offen versus geschlossen – Welchen Zusammenhang gibt es zwischen Apple iOS- und Android-App-Entwicklern? Wirtschaftsinformatik, Onsabrück, Germany (2015)
58. Lee, C., Lee, D., Hwang, J.: Platform Openness and the Productivity of Content Providers: A Meta-Frontier Analysis. Telecommunications Policy 39, 553-562 (2015)
59. van Angeren, J., Alves, C., Jansen, S.: Can We Ask You to Collaborate? Analyzing App Developer Relationships in Commercial Platform Ecosystems. Journal of Systems and Software 113, 430-445 (2016)
60. Gawer, A.: What Drives Shifts in Platform Boundaries? An Organizational Perspective. Academy of Management Proceedings 2015, 13765-13765 (2015)
61. Benedict, M., Schlieter, H., Burwitz, M., Esswein, W.: ISO 11354-2 for the Evaluation of eHealth Platforms. 20th Pacific Asia Conference on Information Systems, Chiayi, Taiwan (2016)
62. Hein, A., Schreieck, M., Wiesche, M., Krcmar, H.: Multiple-Case Analysis on Governance Mechanisms of Multi-Sided Platforms. Multikonferenz der Wirtschaftsinformatik, Ilmenau, Germany (2016)

63. Hofer-Shall, Z.: Working Directly With the Twitter Data Ecosystem. https://web.archive.org/web/20150412145411/https://blog.gnip.com/twitter-data-ecosystem/ (Accessed: 25/03/2018)
64. Instagram: Instagram Platform Update. https://web.archive.org/web/20151119094951/http://developers.instagram.com/post/133424514006/instagram-platform-update (Accessed: 25/03/2018)
65. Trachtenberg, A.: Changes to our Developer Program. https://developer.linkedin.com/blog/posts/2015/developer-program-changes (Accessed: 25/03/2018)
66. van den Bosch, T.: The Rise of the Closed APIs. https://blog.deptagency.com/the-rise-of-the-closed-apis-6bd70a353fd5 (Accessed: 25/03/2018)
67. Atkins, M.H.: The Role of Appropriability in Sustaining Competitive Advantage - An Electronic Auction System Case Study. The Journal of Strategic Information Systems 7, 131-152 (1998)
68. Wareham, J., Fox, P.B., Cano Giner, J.L.: Technology Ecosystem Governance. Organization Science 25, 1195-1215 (2014)
69. Ernkvist, M.: Down Many Times, But Still Playing the Game: Creative Destruction and Industry Crashes in the Early Video Game Industry 1971-1986. In: Gratzer, K., Stiefel, D. (eds.) History of Insolvency and Bankruptcy from an International Perspective, pp. 161-191, Stockholm, Sweden (2008)
70. Soeldner, C., Roth, A., Möslein, K.M.: Opening Embedded Systems for Open Innovation. 19th Pacific Asia Conference on Information Systems, Singapore (2015)
71. Holzer, A., Ondrus, J.: Mobile Application Market: A Developer's Perspective. Telematics and Informatics 28, 22-31 (2011)
72. Guo, J., Bouwman, H.: An Analytical Framework for an M-Payment Ecosystem: A Merchants' Perspective. Telecommunications Policy 40, 147-167 (2016)
73. O'Mahony, S., Karp, R.: From Proprietary to Collective Governance: How Platform Participant Strategies Adapt. Boston University, Boston, MA, USA (2017)
74. Song, J., Choi, H.Y., Baker, J., Bhattacherjee, A.: Mobile Application Development Platform Adoption: A Grounded Theory Investigation. 19th Americas Conference on Information Systems, Chiago, IL, USA (2013)
75. Choia, G., Nam, C., Kim, S.: The Impacts of Mobile Platform Openness on Application Developers' Intention to Continuously Use a Platform: From an Ecosystem Perspective. 28th European Regional Conference of the International Telecommunications Society (ITS): "Competition and Regulation in the Information Age", Passau, Germany (2017)
76. Hagiu, A.: Proprietary vs. Open Two-Sided Platforms and Social Efficiency. Harvard Business School Strategy Unit Working Paper. Harvard Business School, Boston, MA, USA (2006)
77. Jansen, S.: How Quality Attributes of Software Platform Architectures Influence Software Ecosystems. 2013 International Workshop on Ecosystem Architectures, Saint Petersburg, Russia (2013)
78. Robillard, M.P., DeLine, R.: A Field Study of API Learning Obstacles. Empirical Software Engineering 16, 703-732 (2010)
79. Vis, B.: The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N Analyses. Sociological Methods & Research 41, 168-198 (2012)
80. Ragin, C.C.: Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA). Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques, (2009)

81. Dellermann, D., Reck, F.: The Antecedents of Transaction Costs in Digital Ecosystems: A Configurational View on the Interplay of App Architecture and Platform Governance. 50th Hawaii International Conference on System Sciences, Waikoloa, HI, USA (2017)
82. Dellermann, D., Reck, F.: Minimizing Complementors' Risk in Third-Party Innovation: A Qualitative Comparative Analysis (QCA) of Digital Platform Configurations. 50th Hawaii International Conference on System Sciences, Waikoloa, HI, USA (2017)
83. Dellermann, D., Jud, C., Reck, F.: Understanding Platform Loyalty in the Cloud: A Configurational View on ISV's Costs and Benefits. 13th International Conference on Wirtschaftsinformatik, St. Gallen, Switzerland (2017)
84. Stummer, C., Kundisch, D., Decker, R.: Platform Launch Strategies. Business & Information Systems Engineering 60, 167-173 (2018)

Competence, Fashion and the Case of Blockchain

Simon Albrecht[1,2], Gunther Gust[1], Jens Strüker[2], and Dirk Neumann[1]

[1] University of Freiburg, Information Systems Research, Freiburg, Germany
{simon.albrecht,gunther.gust,dirk.neumann}@is.uni-freiburg.de.de
[2] Fresenius University of Applied Sciences, Institute of Energy Economics, Frankfurt, Germany
jens.strueker@hs-fresenius.de

Abstract. Technological innovations that are currently 'in fashion'—such as the blockchain—are subject to inflated expectations regarding their potential. Such fashion phenomena get reinforced by knowledge entrepreneurs, whose discourse highlights organizational problems and offer such innovations as solutions. Managers observing this discourse will either follow a fashion or skeptically await its outcomes. In this research project, we aim to analyze whether managers' IT competence helps them to form realistic expectations about the blockchain technology's potential and subsequently to make thoughtful technology decisions. In our model, we operationalize IT fashion as the extent to which firms' decision makers are exposed to discourse with knowledge entrepreneurs. We hypothesize this construct to have a positive effect on these managers' support for the blockchain technology. We then aim at evaluating how managers' IT competence moderates this effect —which has not been studied before.

Keywords: Blockchain, Technology Competence, IT Fashion, Top Management Support

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

The Digital Platform Otto.de: A Case Study of Growth, Complexity, and Generativity

Daniel Fürstenau[1], Daria Anisimova[1], Dieter Masak[2], Hannes Rothe[1], and Matthias Schulte-Althoff[1]

[1] Freie Universität Berlin, Department of IS, Berlin, Germany
{daniel.fuerstenau,daria.anisimova,hannes.rothe, matthias.schulte-althoff}@fu-berlin.de
[2] plenum AG, Management Consulting, Frankfurt am Main, Germany
{dieter.masak}@plenum.de

Abstract. We analyze the growth, complexity, and generativity of the digital platform Otto.de, a revelatory case of a large German company that has opened up its internal IT platform to outside developers. We find indication for a superlinear growth pattern fueled by external developers and the introduction of microservices as well as the emergence of a structural separation within the platform. Furthermore, our research shows ways to explain the generativity of a digital platform based on the attention and activity received.

Keywords: Digital platform, generativity, complexity, platform evolution

1 Introduction, Related Work, and Purpose

The literature on digital platforms has noted important issues of architectural design determining their long-term "evolvability" and success [1–3]. In this paper, we draw on important contributions which highlight the *modularity* and *generativity* of digital platforms. Modularity conceptualizes digital platforms as being composed of a set of more or less independent modules, which are connected via well-defined interfaces to ensure the platform's core functionalities [1]. This gives rise to combinatorial innovation since peripheral modules can be changed without altering the platform's core [4]. *Generativity* in turn allows for new functions and services to be created on top of the platform. It describes a platform's ability to generate new outputs, structures, or behaviors beyond the creators' original intentions [3, 5, 6] and as such is an important impetus and mechanism of innovation. Together, modularity and generativity facilitate further development opportunities and platform evolution.

However, while previous contributions (such as [1–3, 6]) have expanded our understanding on platform evolution, we currently lack a detailed understanding of platform evolution and generativity especially in the context of companies which transform their internal operations and IT platforms into a platform model.

The purpose of this paper is to give insight into the *growth, complexity, and generativity of the digital platform Otto.de*. In doing so, we give a detailed architectural overview of the platform's main characteristics and development since its inception in 2012. This is interesting since in the course of digital transformation many companies

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

embark on a journey of opening up their internal IT platforms to outside contributors, but there are few examples of companies that have successfully transformed a legacy business model into a fully digitally one, as from catalogue shipping to e-commerce. For Otto, the platform today handles more than 90% of the company's order volume (€2.7 billion; 2 million site visits per day; up to ten orders per second) and is vital for its transformation. Moreover, the platform is interesting because it evolved fully from a single business unit's initiative toward the company's main revenue/value creator. Furthermore, it is one of Germany's largest agile development projects in a traditional enterprise—allowing insight into scaling agile in the large. The initiative started with a team of 100 staff members and has, since then, grown to its current size of 250 employees. These are organized in interdisciplinary development teams across functional areas (search, navigation, product presentation, etc.), as detailed in [7]. Finally, it is noteworthy from an architectural point of view since the platform was restructured using a microservice approach over the past years.

2 Methodological Approach and Conceptual Framework

We employ a case study approach [8]. Our data is the GitHub repository of the platform Otto.de[1]. In this paper, we do not focus on the "multi-sidedness" of the platform towards professional sellers, which had just begun in 2017. Yet, the interesting aspect here is that the platform had been "opened up" to external developers, which moves it away from an internal IT platform to a model where combinatorial innovation becomes possible through the contributions of external developers writing code which is integrated into the platform on the level of particular modules (i.e., repositories). We extracted the entire public repository via a Python script; this covers panel data tracing back from the platform's external opening in 2012 until June 2018 (6.5 years). This "digital trace data" [9] was preprocessed and resulted in a detailed overview of the platform's 59 repositories, 351 Contributors, and 8,733 Commits. Several conversations with leading managers and developers of the platform business unit confirmed the representativeness of the data for the project's development in general. For analyzing platform growth, we used the *cumulated sum of the size* (in KB) as a function of *time* (in month), as well as time series analysis drawing on phase average and linear regression methods. For analyzing complexity, we firstly extracted the bipartite relations between repositories and developers and secondly collapsed this into a one-mode network of inter-repository relations via joint developers. Gephi, an open source network analysis tool, was used for visual analysis and metrics calculation. To analyze generativity, we conceptualized *generativity* in terms of the *number of forks* of a repository. The number of forks captures the number of times external projects / developers re-use a repository for their own project or purpose. It thus illustrates the ability to generate new "output" [5] from a particular input (repository) beyond the intention of the original creators. Correlation analysis was used where we posit that *attention on* the repository (measured in terms of the *number of watchers*) and *activity* on the repository (measured in terms of number of open issues) are associated with generativity. Furthermore, we posit that *generativity* is associated with *popularity* of a

[1] https://github.com/otto-de

repository (measured in terms of number of stars). If a repository perceives a high level of attention and activity this may act as a signal for developers to further increase their own activity here, thereby contributing to its overall generativity and popularity [10].

3 Findings

Results of the growth analysis in Fig. 1A point to a stepwise growth; the sharp increase since 2015 indicates a superlinear pattern. This would correspond to the examples of open source software development presented in the work of Scacchi [11]. A conclusion can be drawn from the shape of the growth curve and the trend figure: the activity of the actors is linked to the creation of new repositories. No repositories were created in 2014 and commits stagnated. In 2015, however, 28 repositories were newly created and the activity increased. The pause in 2014 can be explained by the planning and development of concepts for future development. Regression and phase average analysis pointed to further seasonality effects reaching their lowest point in the summer months.

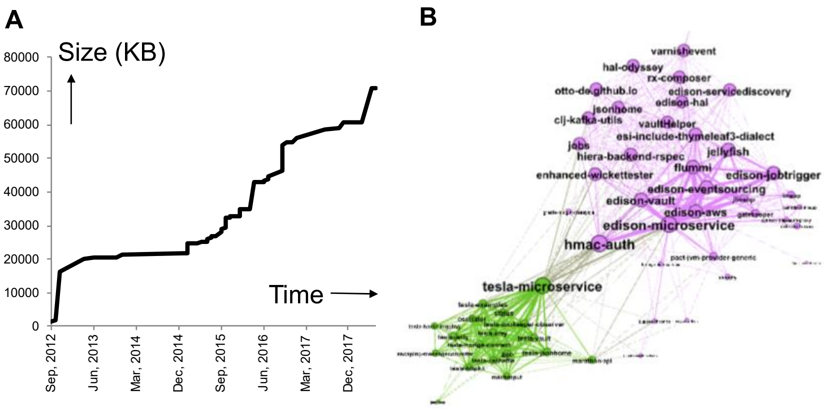


Figure 1. Growth of size in time (a) and network (b) of Otto.de repositories

Complexity analysis pointed to the emergence of two different clusters (see Fig. 1B), which were demarcated by two main hubs (i.e., Tesla and Edison Microservice). This is in consonance with the findings of Singh et al. [12] and Um et al. [13] that the development of a network has a basic structural pattern. The distributed architecture of the digital platform allows the individual "building blocks" to be developed separately. Platform microservices (e.g., Tesla and Edison) are written in different programming languages and controlled by different teams, but adhere to standards on RESTful APIs in order to ease the process of transferring data. The relationship mapping also allows conclusions about the generativity and the knowledge exchange in such externally-infused projects, suggesting that more than 100 third-party developers participated.

Further analyses regarding *generativity* using Spearman's Correlation Coefficient suggested robust to strong associations between *attention* (watchers) and *generativity* ($r(57)=.29$, $p<0.05$), *activity* (open issues) and *generativity* ($r(57)=.35$, $p<0.01$), as well as *generativity* and *popularity* (stars) ($r(57)= .68$, $p<0.001$). Further research aims at developing a multivariate model for *explaining* generativity including controls such as age or quality of a repository. Moreover, it would be interesting to link generativity to user base-related popularity measures such as platform usage or user satisfaction.

4 Concluding Remarks

The study has preliminary implications for academics interested in platform evolution and companies wishing to open their internal IT platforms. First, modularitization of the platform enables superlinear growth, but also creates the possibility that structural patterns and separations emerge which are reinforced over time. Second, retaining outside contributions from external developers over longer time periods seems possible, but we need a greater understanding of the mechanisms of disembedding platforms from local contexts and developer participation in mixed corporate-volunteer contexts. Further research should thus investigate in more detail how the e-commerce company and others collaborate online and offline with different developer and stakeholder groups working on different components of a digital platform.

References

1. Tiwana, A., Konsynski, B., Bush, A.A.: Research Commentary - Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. Inf. Syst. Res. 21, 675–687 (2010).
2. Agarwal, R., Tiwana, A.: Editorial—Evolvable Systems: Through the Looking Glass of IS. Inf. Syst. Res. 26, 473–479 (2015).
3. de Reuver, M., Sørensen, C., Basole, R.: The Digital Platform: A Research Agenda. J. Inf. Technol. 1–12 (2017).
4. Yoo, Y., Henfridsson, O., Lyytinen, K.: Research Commentary - The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. Inf. Syst. Res. 21, 724–735 (2010).
5. Zittrain, J.: The Future of the Internet - And How to Stop It. Yale University Press, London (2008).
6. Henfridsson, O., Bygstad, B.: The Generative Mechanisms of Digital Infrastructure Evolution. MIS Q. 37, 907–931 (2013).
7. Kraus, S., Wolter, P.: Scaling Agile @ Otto - Learning at Scale: Agilität als organisatorische Herausforderung. OBJEKTspektrum. 20–24 (2016).
8. Yin, R.K.: Case Study Research: Design and Methods. Sage Publications, Thousand Oaks, California (2013).
9. Eck, A., Uebernickel, F.: Reconstructing Open Source Software Ecosystems: Finding Structure in Digital Traces. In: ICIS 2016 Proceedings (2016).
10. Dabbish, L., Stuart, C., Tsay, J., Herbsleb, J.: Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. In: Proceedings of the ACM 2012 Conference on CSCW. pp. 1277–1286 (2012).
11. Scacchi, W.: Understanding Open Source Software Evolution. In: Madhavji, N.H., Fernández-Ramil, J.C., and Perry, D.E. (eds.) Software Evolution and Feedback: Theory and Practice. Wiley, New York.
12. Singh, P.V., Tan, Y., Mookerjee, V.: Network Effects: the Influence of Structural Capital on Open Source Project Success. Mis Q. 35, 813–829 (2011).
13. Um, S., Yoo, Y., Wattal, S.: The Evolution of Digital Ecosystems: A Case of WordPress from 2004 to 2014. In: ICIS 2015 Proceedings (2015).

Track 8:

eHealth & alternde Gesellschaft

Track Chairs:

Prof. Dr. Claudia Müller
Universität Siegen

Prof. Dr.-Ing. Michael Prilla
TU Clausthal

Security and Privacy of Personal Health Records in Cloud Computing Environments – An Experimental Exploration of the Impact of Storage Solutions and Data Breaches

Michael Adelmeyer[1], Pascal Meier[1], and Frank Teuteberg[1]

[1] University of Osnabrück, Institute of Accounting and Information Systems, Germany
{michael.adelmeyer,pascal.meier,frank.teuteberg}@uos.de

Abstract. In the course of the digitization in healthcare, the collection and central storage of large health-related datasets in clouds in the form of personal health records is growing. However, the use of cloud services for sensitive data is associated with security and privacy risks. Further, the delegation of control over security and privacy measures to the cloud provider requires trust on the users' side. In order to investigate the role of security and privacy when storing and processing patient data, we conducted an online experiment, in which third-party cloud services are compared to private on-premise data centers. Additionally, we examine the impact of data breaches on the perceived security, privacy, control and trust in both storage scenarios. Our results indicate that cloud-based personal health records still face concerns regarding perceived security, privacy, control and trust amongst end-users. Nevertheless, after a data breach, no significant differences between both solutions exist.

Keywords: Cloud Computing, Personal Health Records, Security, Privacy.

1 Introduction

Personal health records (PHRs) can improve the collaboration between different healthcare actors and patients and ease the sharing of data [1] by providing central repositories of the users' state of health [2]. When providing PHRs for patients, cloud computing (CC) offers numerous benefits, such as scalable computing resources at minimal costs [3]. Yet, adopting CC services for sensitive health data introduces risks associated to security and privacy, as the data are stored externally or even in public environments, which entails additional challenges, such as access control or privacy protection [4], [5]. Compared to private on-premise data storage, the control over the data and corresponding security and privacy measures is delegated to the cloud provider, which requires trust in the capabilities of the provider on the users' side [4]. Since data security and privacy are crucial determinants for the adoption of PHRs [6], [7], it is necessary to evaluate whether patients' data should be stored and processed on-premise or externally, via scalable CC services. For PHR providers, the users'

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

acceptance, which is prerequisite for a sustainable adoption of the provided service, heavily depends on the selection of a secure and trustworthy storage solution [8].

Generally, data security and privacy issues are regarded as a central inhibitor for the adoption of cloud-based PHRs [5], [8]. For example, the cloud-based PHR app *vivy*, which was recently launched by several major German health insurances, faces massive criticism for data security and privacy issues, e.g., related to data hosting and processing at the underlying Amazon cloud platform [9]. Due to the high value and sensitivity of the data stored in PHRs, cloud environments are an attractive target of malicious attacks [5]. In case of a data breach, the security or privacy of the protected health data are compromised due to an impermissible use or disclosure [10]. The delegation of control in cloud environments [5], [11] and the corresponding need for trust [4] are connected to the perceived security and privacy of PHR users and in view of the theory of reasoned action [12], ultimately influencing the intention to use or adopt cloud-based PHRs [6]. Since CC faces skepticism [11], it is necessary to determine the importance of the factors that influence the intention to use cloud-based PHRs services to ensure their sustainable adoption. However, previous research either focuses on security or privacy related aspects [1], [5], [11], [13–15] or lacks a focus on either PHRs [11], [16] or CC [6], [7], [17]. Although data breaches continually impact the healthcare sector [5], [10], little research has been conducted on their outcome on users of cloud-based PHRs. Therefore, we adopt an experimental approach to investigate the participants' perceived security, perceived privacy, trust, perceived control and intention to use when using a PHR of a provider who stores and processes personal health data (a) at a cloud provider or (b) an on-premise data center. Additionally, the impact of data breaches on the targeted constructs is investigated. We examine the following research questions (RQ):

RQ1: Do patients' perceived security, perceived privacy, trust and perceived control differ between cloud environments and private on-premise data centers when storing personal health data in PHRs?

RQ2: In case of a data breach, to which extent do patients' perceived security, perceived privacy, trust and perceived control differ when using cloud environments compared to private on-premise data centers?

Based on the RQs, the study is structured as follows: First, hypotheses (H) targeting the above-mentioned constructs are derived from the literature. Second, the experimental approach and design are described. Afterwards, we analyze the collected data. In the conclusion, the findings are discussed.

2 Theoretical Background and Hypotheses Development

2.1 Data Security and Privacy

Although CC plays a significant role in healthcare, the technology still faces data security and privacy concerns, especially in PHR provisioning [8], [14], [15], [17], [18]. As the two concepts are strongly connected, individuals often may not fully understand the difference between security of systems and privacy of their data [19],

[20]. Individuals choose to disclose information, although this might impact their privacy [21], which is decisively based on trust and the perception of risks related to data security and privacy. Yet, individuals may decide to refrain from disclosures that would affect the security of their data [19]. Therefore, it is necessary to examine both concepts individually [19], which is not always the case in existing literature on cloud-based PHRs [13], [14]. Although multiple definitions exist, we adapt the most common-sense understanding of privacy [22] as an individual's ability to control the use of his or her own data [13], [20]. Besides perceived control [11], trust is a central determinant of users' intention to disclose information [21]. Whereas privacy focuses on individuals' control over their data, security is an overarching principle with focus on data confidentiality, integrity and availability [17]. Hence, in order to comply with privacy requirements, data security provides the technical foundations [18], [20].

Since certain information stored in PHRs is shared with multiple actors, risks for data security and privacy arise [15], [17]. Consequently, the infrastructure via which PHRs are provided should be rigorously secured [8], [17]. However, this is a central challenge for cloud environments, especially for public clouds [4], [14], [18], where the probability of unauthorized access to data is significantly higher compared to traditional systems [23]. Therefore, data security is of great importance for the provision of PHRs via cloud environments [1]. Hence, on the side of PHR and cloud service providers, extensive data security and privacy requirements must be considered to guarantee an equal level of data security compared to on-premise data centers [8]. Thus, we hypothesize that users' perceptions of security (SEC) and privacy (PRI) are more positive when PHRs are stored and processed on-premise, as using cloud services introduces additional risks and uncertainties [1], [4], [13], [24]:

H1a/H2a: The users' SEC/PRI is higher when their data are stored in on-premise data centers compared to cloud environments.

Despite the sensitivity of PHR data, the risks associated to CC and the higher probability of data disclosures in cloud environments [23], the impact of actual data breaches on users' SEC and PRI is only scarcely explored. If the security and privacy of users' personal health data were compromised in a data breach, this will likely lead to increased perceived risks and an increased motivation for protection [19]. Due to the value and sensitivity of health data stored in PHRs, data breaches occur frequently [10] and potential consequences are severe [13], regardless of the underlying storage solution. From the service users' perspective, the impact of and dealing with potential consequences are the same irrespective of whether a data breach occurred within a cloud or a conventional IT environment [23]. Thus, we hypothesize that the individuals' levels of SEC and PRI do not differ after a data breach:

H1b/H2b: After a data breach, the users' SEC/PRI is equal regardless of whether the data are stored in on-premise data centers or cloud environments.

2.2 Trust

Trust (TR) is defined as the willingness of a trustor to be vulnerable to the actions of a trustee, based on the expectation that the trustee will perform a particular action relevant to the trustor, irrespective of the ability to monitor or control the trustee [25]. In the context of the acceptance of information systems, different trust targets exist, such as trust in a technology or a provider [26]. With regard to PHRs, the user takes the role of the trustor and the PHR provider the role of the trustee, who may utilize CC as underlying technology [24]. For the general adoption and use of cloud services, trust has been identified as a key factor [24], [27], as users delegate their control over data security and privacy to the provider, whose actions, in turn, are difficult to monitor [4]. In healthcare, users' trust in security and privacy mechanisms of cloud providers is seen as a core determinant [8], [13]. The trustor must be confident that the cloud technology adopted in a certain healthcare data sharing scenario complies with security and privacy requirements [18]. Hence, when storing and processing personal health data in a PHR, we conclude that users' trust in the PHR provider is lower when the data are stored in a cloud environment compared to on-premise data centers directly operated by the provider, since users perceive more risks associated to CC [11], [13], [16]. However, there should be no difference in the users' trust after a data breach, as the perceived risks and the corresponding lower trust levels towards the technology are superposed by the decrease of trust in the PHR provider [26]:

H3a: The users' TR is higher when their data are stored in on-premise data centers compared to cloud environments.
H3b: After a data breach, the users' TR is equal regardless of whether the data are stored in on-premise data centers or cloud environments.

Although most studies investigate the influence of perceived risks on trust [20], Mitchell (1999) identified that there is a bi-directional relationship between TR and perceived risks [28] that include SEC and PRI. This has, among others, been empirically proven by Hsin Chang and Wen Chen (2008) [29]. Regarding the acceptance of health clouds, trust in healthcare cloud providers and in privacy-preserving mechanisms influence the patients' information privacy concerns [13]. Thus, focusing on the role of security and privacy when storing personal health data in PHRs, we analyze the influence of TR on SEC and PRI:

H4/H5: The higher the users' TR, the higher the SEC/PRI.

2.3 Control

Another important aspect of privacy is the perceived control (CO) over a system or technology [30], which is largely determined by the extent to which a user is able to employ a system in order to achieve an intended goal [31]. In cloud environments, the users' control over data and privacy or security measures is largely delegated to the cloud provider [11], [24]. In consequence, users perceive lower control over their data, which is why they connect CC with insecurities [11]. Thus, CO additionally refers to the extent to which users perceive that they have an impact on the control over their activities and data in the cloud service used [32]. Due to the sensitivity of data stored in a PHR, it is necessary that the users perceive to be in control and to be able to decide over the access to and the usage of their data. In this context, CO aims at both internal and external control. This means that the user feels capable of using the system and further, that the situation allows this [33]. Compared to on-premise data centers, the provision of PHRs via cloud environments mainly differs in terms of external control, since CC is afflicted with uncertainties regarding the security and privacy of data [4], [24]. However, in case of a data breach occurring at either storage solution, we hypothesize that the awareness of uncertainties and risks connected to CC services only plays a minor role, as a loss of control over information is regarded as material for the perception of a privacy invasion [11]. Consequently, the CO for both PHR data storing and processing solutions adjusts to a comparably low level:

H6a: The users' CO is higher when their data are stored in on-premise data centers compared to cloud environments.

H6b: After a data breach, the users' CO is equal regardless of whether the data are stored in on-premise data centers or cloud environments.

As trust is defined irrespectively of the ability to monitor or control the trustee [25], we do not assume a connection between CO and TR. Yet, SEC and PRI of PHRs are often linked to users' CO, for example, by implementing access controls for PHRs [17] or health data in clouds in order to increase data security and privacy [15]. In cloud environments, recent studies suggest that cloud users' CO majorly determines their PRI [11]. Hence, fostering users' CO by means of security and privacy measures enhances the users' SEC and PRI when storing and processing health data in PHRs:

H7/H8: The higher the users' CO, the higher the SEC/PRI.

2.4 Intention to Use

Following the theory of reasoned action, individuals' intentional behaviors to adopt or use a system are ultimately determined by their attitudes [12], [34]. The intention to use (ITU) can be defined as the intention of a customer or service user to (continually) use a provider and his service (also after an incident) [24], [32]. Besides a general attitude towards a service, the ITU of a person is measured based on risks and benefits regarding a scenario or technology [13], [24]. Due to the steadily increasing complexity of technology, trust plays a crucial role in the acceptance of and the ITU information systems [26], such as cloud environments. In view of the impact of users' security and privacy concerns towards CC, higher levels of security and privacy perceptions positively affect the ITU cloud services [16]. In this context, PRI and SEC are majorly determined by the delegation and the corresponding loss of control in cloud environments [11], [18], which ultimately affect the ITU cloud services [11]. In healthcare, control, trust, security and privacy are proven to influence the ITU cloud services, as end-users are not able to fully control the storage and processing of their data due to the control delegation and the opaque nature of cloud services [18], [24] as well as due to the sensitivity of data [13], [18]. Especially security and privacy are regarded as important requirements for the adoption of cloud-based PHRs [6], [8]. Further, trust is an antecedent of individuals' disclosure behavior [21]. Thus, it is assumed that SEC, TR, CO and PRI have a positive impact on the ITU in PHR data sharing scenarios:

H9/H10/H11/H12: The higher the users' SEC/TR/CO/PRI, the higher the ITU.

2.5 Research Model

The constructs and hypotheses are summarized in a research model (see Figure 1).

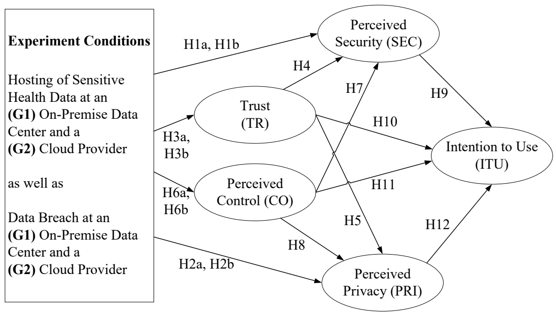


Figure 1. Research Model

3 Research Approach

3.1 Experimental Design and Procedure

In order to test the previously developed hypotheses, we conducted a web-based experiment using vignettes [35]. Overall, we applied a two-group posttest-only design [36], which was extended by an additional stimulus (X_2) and a second posttest (P_2) in both groups to determine the impact of a data breach on the outlined constructs. The overall procedure of the experiment is visualized in Figure 2.

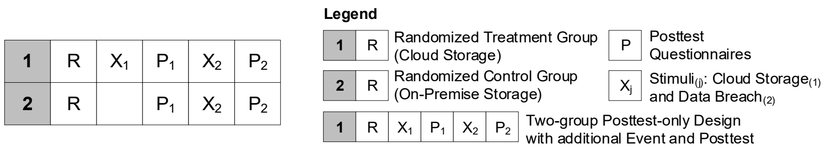


Figure 2. Experimental Setup

The random allocation (R) of the participants to the two experiment groups G1 and G2 took place in an unnoticeable first step of the experiment. Afterwards, a vignette was displayed, which can be described as a focused description of a hypothetical scenario that puts the subjects in a fictitious role in which they are asked to respond in certain ways [35]. Our vignette[1] introduced the participants to the German healthcare system, in which health-related data are stored decentrally. As a result, a medical treatment based on all existing health data is difficult, because stakeholders often cannot access all patient data. The participants are asked to act as longtime users of a service called *Vitalife*, which offers a PHR web service. The service provides a central storage and management of users' relevant personal health and related data. The first

[1] The full scenarios can be retrieved under the following link: http://bit.ly/WI-19

manipulation between group 1 (G1, treatment) and group 2 (G2, control) was conducted by using two data storage solutions of the PHR service provider *Vitalife*, i.e., a third-party cloud service (G1) (X_1) and a private on-premise data center directly operated by *Vitalife* (G2). The participants of both groups were identically informed about data security and privacy measures of the respective systems in use. Both storage solutions comply with common information security standards, such as ISO 27001, which require the development and implementation of a strict security program. Further, the third-party cloud service used by *Vitalife* complies with the cloud-related ISO 27017/27018 standards for IT security controls and the processing of personal data in cloud services. Both solutions offer an availability of 99.9 percent.

After the vignette, the first posttest (P_1) is conducted to measure the participants' attitudes towards the different storage solutions prior to a data breach. In the subsequent textual description of the event [24], [27] (X_2), the subjects are informed by *Vitalife's* customer service about a data breach at the employed cloud service provider (G1) or the on-premise data center (G2), in the course of which the PHR data of the participants were stolen. In both groups, an employee of *Vitalife* has been made responsible for the data breach, as he inadvertently opened a link in a phishing e-mail and thus enabled hackers to access the system. In order to allow for a comparison of the users' SEC, PRI, CO and TR before and after the data breach, a second identical posttest (P_2) was conducted directly after the event in both groups. In both tests, the participants were ultimately asked for their ITU the service based on their experience.

3.2 Operationalization of the Constructs

In the questionnaires, the subjects were surveyed regarding the outlined hypothetical constructs using several items derived from the literature (see Table 1).

Table 1. Constructs and Item Sources

Construct	*Adapted Definition*	*Source*
SEC	The perception of PHR service users regarding the capabilities of a provider to protect their data, which include safe transmission, prevention of unauthorized modification, access or interference.	[20]
PRI	The PHR service users' perceived ability to control the acquisition and use of their personal information [20], which includes the general concern of the provider, the adherence to privacy laws or the prevention of unauthorized information disclosure.	[20]
TR	The willingness of a PHR service user to be vulnerable to the actions of a PHR service provider, based on the expectation that the provider safely stores and provides patients' health data [25].	[37]
CO	The extent to which PHR service users perceive that they can impact the storing and processing of their health data as well as the access rights of other stakeholders to their data.	[31]
ITU	The intention to voluntarily use the provided PHR service, which stores health data in the cloud or in an on-premise data center.	[13], [32]

All items were measured by means of a seven-point Likert scale ranging from 1 ("strongly agree") to 7 ("strongly disagree"). For SEC and PRI, 6 items each from Flavián and Guinalíu (2006) [20] were used. TR was measured based on 11 items from McKnight et al. (2002) [37], with 4 items targeting the integrity (TRI) and ability (TRA) each as well as 3 items measuring the benevolence (TRB) of *Vitalife*. For CO, 4 items of Lee and Benbasat (2011) [31] were used. ITU was measured with 3 items adapted from Ermakova (2014) [13] and one from Walter et. al (2014) [32].

3.3 Participants and Data Collection

To test our proposed hypotheses, the previously described experimental concept was carried out as a web-based survey in February 2018. In total, 238 undergraduate economics and information systems students participated in the voluntary experiment, which was rewarded with minor raffled incentives. Out of the 238 answers, we only considered complete datasets for the analysis and eliminated answers with response patterns or significantly faster overall response times. Among the remaining 217 datasets, 104 participants belong to group 1 and 113 to group 2, with an overall of 35 % female and 65 % male participants and an average age of 20.6 years.

4 Data Analysis and Results

4.1 Construct Validity and Reliability

In order to ensure the validity and reliability of the collected data, several tests were conducted. The analysis of variances tests (ANOVA), Levene tests of variance equality and t-tests for the control variables gender, age, health condition and CC experience did not indicate significant differences between the groups (N = 217). Therefore, a homogeneous distribution of the participants can be assumed. Given a mean self-estimation of 3.37, it can be concluded that the participants are fairly knowledgeable in CC. Since most of the hypotheses and the research model are based on the first posttest, the validity and reliability are assessed based on the corresponding data and are thus concluded for the second posttest. The data were examined in order to preclude a common method bias (CMB) using Harman's one-factor test [38]. The 30 indicators used were examined in a factor analysis. Since the resulting extraction of one factor explaining 45.783% of the variance is below the threshold of 50%, we do not assume a CMB [38]. Subsequently, we assessed the one-dimensionality of the constructs in an exploratory factor analysis. All indicators meet the respective thresholds for Measure of Sampling Adequacy (MSA) (> .5), the Kaiser-Meyer-Olkin-criterion (KMO) (> .5), communalities (> .5) and the Bartlett-test (< .05) [39], except from the items TRB2 and TRB3 with communalities < .5 (.450 and .498). Hence, both items were dropped from further analyses, which, using reflective measurement, does not affect the meaning of a construct [39], [40]. Further, according to Kaiser's rule, five common factors with an eigenvalue of > 1, which explain 66.915% of the variance, could be extracted. This confirms our predetermined five constructs SEC, PRI, CO, TR and ITU [39]. Although items with loadings > .6 are considered as reliable [41], we dropped all items with loadings < .7 in order to achieve better construct validity, as suggested by Nunnally and Bernstein (1994) [42]. This concerns our items TRA1 (.699), TRA4 (.576), PRI4 (.698) and SEC2 (.680). Further, since the cross loadings of the items PRI5 (.697 to TR), TRI1, TRI3 and TRI4 (.643, .657 and .692 to PRI) are relatively high, the respective items were dropped in order to strengthen discriminant validity, which did not adversely affect our previously conducted analyses.

The construct's internal consistency reliability was assessed calculating the Cronbach's Alpha (CA), the inter-item correlation (IIC) and the corrected inter-scale correlation (CISC). For CA, IIC and CISC, the recommended threshold values (CA ≥ .7; IIC ≥ .3; CISC ≥ .5) are met for all constructs (see Table 2). The reliability of the constructs in terms of quality criteria of second order was assessed examining the composite reliability (CR) and the average variances extracted (AVE) (see Table 2). For both measures, the thresholds are met (CR ≥ .6; AVE ≥ .5). Regarding the indicator reliability (IR) (not listed in Table 2), the required threshold of ≥ .4 is met by all indicators (lowest IR = .545). Therefore, we assume convergent validity in the context of construct validity. Next, we evaluated the discriminant validity by checking for the Fornell-Larcker criterion, which states that the square roots of the AVE (dark cells) need to be higher than the corresponding correlations between the constructs [43]. Since this is the case for all our constructs, discriminant validity for all factors can be assumed. Thus, our constructs serve for further hypotheses testing.

Table 2. Discriminant Validity

Factor	*# Items*	*CISC*	*IIC*	*CA*	*CR*	*AVE*	*CO*	*ITU*	*PRI*	*SEC*	*TR*
CO	3	.754-.829	.752	.901	.938	.835	.914				
ITU	4	.701-.856	.708	.906	.934	.781	.578	.884			
PRI	4	.623-.683	.539	.824	.883	.653	.629	.629	.808		
SEC	5	.682-.830	.680	.914	.936	.746	.592	.607	.640	.863	
TR	4	.527-.654	.486	.791	.865	.616	.555	.580	.681	.577	.785

4.2 Hypotheses Testing

To examine the influence of the experiment conditions on the latent variables TR, CO as well as on SEC and PRI, we conducted analyses of means (two-sample t-test). To evaluate the differences between the constructs regarding cloud service (G1) and on-premise data storage (G2) before ("a" hypotheses, G1 and G2) and after a data breach ("b" hypotheses, G1E and G2E), the results of the first and the second posttest were compared between G1 and G2 (see Table 3). As the Levene tests are not significant, we assume variance homogeneity for all constructs. Since significant differences for all constructs between G1 and G2 ($p < .05$) were identified, H1a, H2a, H3a and H6a are supported. Regarding the outcome of the event on the hypotheses, no significant differences can be identified for SEC, PRI and TR except for CO ($p = .026$). Hence, except from H6b, we find support for H1b, H2b and H3b.

Table 3. Significance Levels and Mean Values

Construct	*Result*	*G1 and G2*	*G1E and G2E*
SEC	Mean Value G1; G2	3.217; .2.890	4.931; 4.850
	Sig. t-Test (Levene)	.038 (.116)	.617 (.206)
PRI	Mean Value G1; G2	2.623; 2.352	3.421; 3.188
	Sig. t-Test (Levene)	.033 (.735)	.091 (.055)
TR	Mean Value G1; G2	2.618; 2.316	3.447; 3.232
	Sig. t-Test (Levene)	.003 (.958)	.114 (.108)
CO	Mean Value G1; G2	2.798; 2.443	3.683; 3.295
	Sig. t-Test (Levene)	.029 (.506)	.026 (.388)

The relationships between the constructs and their impact on the dependent variables were evaluated using partial least squares (PLS) structural equation modeling (SEM) with SmartPLS 3.0 [44]. In order to evaluate the model fit, we checked for the CMIN/DF as well as the SRMR of the model (2.661, .073), which meet the required thresholds (CMIN/DF $\leq$ 3.0, SRMR $\leq$ .08) [45], [46]. Since further fitness criteria range slightly above or below the recommended thresholds (in brackets), i.e., CFI = .909 ($\geq$ .9), RMSEA = .088 ($\leq$ .08, should not exceed $\geq$ .1) and GFI = .837 ($\geq$ .9), we assume a good model fit. Especially RMSEA values are debatable and thus drawing absolute cutoffs is inadvisable [41], particularly when considering smaller sample sizes with $n \leq 250$ [46]. Afterwards, the path correlations of our SEM were evaluated in order to identify significant correlations between the constructs (see Figure 3).

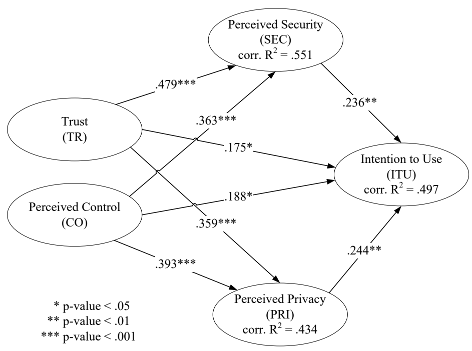


Figure 3. PLS Model

Strong statistically significant relationships between TR and SEC and PRI as well as between CO and SEC and PRI can be confirmed. Hence, we find support for H4, H5, H7 and H8 in our model. Further, path coefficients of $\geq$ .236 between SEC as well as PRI and ITU point to a moderate influence of both constructs on ITU. Therefore, H9 and H10 can be confirmed. Since CO and TR largely determine SEC and PRI (R^2 = .551 and .434), which again have a strong effect on the ITU and despite the relatively low but still significant ($p < .05$) path correlations (.175 and .188), we assume a low but significant impact of TR and CO on ITU. Thus, H10 and H11 can be confirmed. Based on the analyses, the statuses of our hypotheses are shown in Table 4.

Table 4. Statuses of the Hypotheses

H	*Relation*	*Status*	*H*	*Relation*	*Status*
1a	SEC(G2) > SEC(G1)	Supported	6a	CO(G2) > CO(G1)	Supported
1b	SEC(G2E) = SEC(G1E)	Supported	6b	CO(G2E) = CO(G1E)	Not Supported
2a	PRI(G2) > PRI(G1)	Supported	7	CO ↑ ⇒ SEC ↑	Supported
2b	PRI(G2E) = PRI(G1E)	Supported	8	CO ↑ ⇒ PRI ↑	Supported
3a	TR(G2) > TR(G1)	Supported	9	SEC ↑ ⇒ ITU ↑	Supported
3b	TR(G2E) = TR(G1E)	Supported	10	TR ↑ ⇒ ITU ↑	Supported
4	TR ↑ ⇒ SEC ↑	Supported	11	CO ↑ ⇒ ITU ↑	Supported
5	TR ↑ ⇒ PRI ↑	Supported	12	PRI ↑ ⇒ ITU ↑	Supported

5 Conclusion

5.1 Discussion and Implications

The results of the structural equation model indicate that TR, CO, SEC and PRI are key determinants of the intention to use PHR services. In addition, TR and CO both significantly determine SEC and PRI (H4, H5, H7, H8), which ultimately determine the ITU the provided service, independently of the underlying storage solution (cloud vs. on-premise) (H9, H10, H11, H12). Given the similar outcome of both - the storage solution and the data breach - on SEC and PRI, it is possible that the participants did not significantly differentiate between the constructs [19]. Thus, apart from the necessity to comply with strict healthcare regulations, PHR providers need to consider measures that strengthen the TR, CO and especially the SEC and PRI of potential users, e.g., by providing transparency over control, security and privacy measures. Although previous research found that the positive aspects of health clouds outweigh patients' privacy concerns [13], the results of the experiment reveal that the subjects' SEC, PRI, CO and TR are significantly higher when personal data in the form of PHRs are stored on-premise compared to cloud environments (H1a, H2a, H3a, H6a) (RQ1). Thus, despite the relatively high maturity of the technology and the efforts undertaken by theory and practice, CC still faces skepticism regarding the provisioning of PHRs. Providers of PHR solutions need to be aware of this perception and have to decide whether to offer their services via cloud environments or on-premise solutions. When offering cloud-based PHR services, providers might rely on private clouds in order to minimize concerns on the users' side. Further, hybrid cloud infrastructures can be used for encrypted or non-sensitive data in order to realize cloud-related benefits. However, in case of a data breach, no significant differences regarding SEC, PRI and TR between the two solutions could be determined (H1b, H2b, H3b), except for CO (H6b) (RQ2). In cloud environments, data breaches might entail a considerably larger impact on the side of the provider compared to on-premise environments [23]. Hence, cloud services need to emphasize their measures regarding security and privacy and foster the trust of (potential) end-users, e.g., by providing certifications [8], as cloud users seek for external control agents [11]. In order to avoid severe consequences for PHR service users in case of a data breach, strong encryption mechanisms should be implemented [1], [15]. Although both storage solutions were compliant to ISO27001, the SEC, PRI and TR in the cloud scenario were found to be lower. Thus, besides providing certifications, further measures are necessary in order to foster users' perceptions of the constructs. In addition, the experiment revealed that the participants' CO regarding the on-premise solution is still significantly higher after a data breach, even though the participants' abilities regarding control did not vary between the experiment groups. This is presumed to be ascribed to the control delegation and the associated reservations towards cloud environments, which draws the attention of cloud-based PHR providers on establishing measures to particularly foster their users' CO.

5.2 Limitations and Future Research

Since the applied vignette technique puts the participants in a fictitious scenario, the behavior of the subjects might be different compared to a real-world scenario [35]. In addition, using students as experiment participants is not without controversy in the field of information systems and partially limits the generalizability of the results. However, since students reflect the future target group and are early adopters of innovative technologies, such as CC and PHRs, we argue that students are an adequate sample group for the experiment conducted [24], [32]. In this context, it is necessary to note that the experiment was conducted based on a relatively homogeneous sample group, which might influence the overall results. Concerning the evaluation of the model fit, our research model revealed slight weaknesses in terms of some fit indices. Nevertheless, since the deviations from the recommended thresholds are relatively small and cutoff thresholds are discussed controversially [41], we argue that the overall model is not endangered. Further, although established items for measuring trust were derived from the literature [37], several items had to be dropped. Yet, since we used reflective measurement, the meaning of the constructs TR, PRI and SEC was not affected [39], [40]. However, using constructs which strongly differ in the number of items (e.g., TR and ITU) might lead to distorted results. In addition, due to the temporal proximity of the two posttests conducted, an impact of the first on the second survey resulting in a certain bias is possible.

Regarding the data breach, the manipulation of the subjects can possibly be ascribed to multiple factors. In G1, the data are stored externally at a cloud provider, whereas in G2, the data are stored on-premise, which might lead to different accusations towards the actors. However, the examined scenarios reflect the practical problem of companies whether to provide their services via external cloud solutions or on-premise data centers [8], [17]. In this context, it needs to be noted that the use of certain PHR systems might not be purely voluntarily, but instead stipulated by politics or certain healthcare actors. Although control is seen as a core principle of privacy [30], the adoption of privacy as control in the course of this study can be regarded as simplification of the complex nature of the construct [22], [30]. As we concentrated on trust and control as influencing factors of SEC and PRI, the examination of further constructs or the application of behavioral explanation models is possible.

Our study provides multiple starting points for future work. For example, besides investigating TR and CO as influencing factors of PRI and SEC, further constructs (e.g., benefits) can be examined regarding their influence on the two constructs. Additionally, the cause for the higher perceived control in the on-premise solution after a data breach and the correspondingly adjusted levels of SEC, PRI and TR allow for further investigation. A practical validation of the results by means of an examination of a commercial PHR provider, e.g., in the form of a case study, is desirable. In this context, the implications and results from the present study may serve as a foundation for additional qualitative studies or the examination of actual privacy behaviors. Moreover, our experimental setup even allows for richer insights and analyses, e.g., a cross-comparison between the groups before and after the breach.

6 Acknowledgements

The authors would like to thank the experiment participants, the other project members, specifically Ms. Miftari and Ms. Imhorst, for their valuable and substantive help, as well as the reviewers for their constructive feedback. This contribution was prepared within the research project “Dorfgemeinschaft 2.0”. The project is funded by the Federal Ministry for Education and Research (BMBF) (funding code 16SV7453).

References

1. Kaletsch, A., Sunyaev, A.: Privacy Engineering: Personal Health Records in Cloud Computing Environments. In: Proceedings of the 32nd International Conference on Information Systems. Shanghai, China (2011).
2. Menachemi, N., Collum, T.H.: Benefits and Drawbacks of Electronic Health Record Systems. Risk Management and Healthcare Policy. 4, 47–55 (2011).
3. Kuo, A.M.-H.: Opportunities and Challenges of Cloud Computing to Improve Health Care Services. Journal of Medical Internet Research. 13 (2011).
4. Zissis, D., Lekkas, D.: Addressing Cloud Computing Security Issues. Future Generation Computer Systems. 28, 583–592 (2012).
5. Li, M., Yu, S., Zheng, Y., Ren, K., Lou, W.: Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption. IEEE Transactions on Parallel and Distributed Systems. 24, 131–143 (2013).
6. Muhammed, I., Wickramasinghe, N.: User Perceptions and Expectations of the Personally Controlled Electronic Health Record (PCEHR): A Case Study of Australia’s e-Health Solution. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 3441–3450. HI, USA (2017).
7. Angst, C.M., Agarwal, R.: Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quarterly. 33, 339–370 (2009).
8. Rodrigues, J.J.P.C., De La Torre, I., Fernández, G., López-Coronado, M.: Analysis of the Security and Privacy Requirements of Cloud-Based Electronic Health Records Systems. Journal of Medical Internet Research. 15 (2013).
9. Tschirsich, M., Schröder, T.: Schwachstellen in Gesundheits-App Vivy, https://www.modzero.ch/static/vivy-app-security-final.pdf (Accessed: 20.11.2018)

10. Wikina, S.B.: What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches. Perspectives in Health Information Management. 11, (2014).
11. Lang, M., Wiesche, M., Krcmar, H.: Perceived Control and Privacy in a Professional Cloud Environment. In: Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 3668–3677. HI, USA (2018).
12. Fishbein, M., Ayzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975).
13. Ermakova, T., Fabian, B., Zarnekow, R.: Acceptance of Health Clouds - a Privacy Calculus Perspective. In: Proceedings of the 22nd European Conference on Information Systems. Tel Aviv, Israel (2014).
14. Plachkinova, M., Alluhaidan, A., Chatterjee, S.: Health Records on the Cloud: A Security Framework. In: International Conference on Health Informatics and Medical Systems, pp. 152–158. Las Vegas, NV, USA (2015).
15. Li, M., Yu, S., Ren, K., Lou, W.: Securing Personal Health Records in Cloud Computing: Patient-Centric and Fine-Grained Data Access Control in Multi-Owner Settings. In: International Conference on Security and Privacy in Communication Systems, pp. 89–106. Singapore (2010).
16. Arpaci, I., Kilicer, K., Bardakci, S.: Effects of Security and Privacy Concerns on Educational Use of Cloud Services. Computers in Human Behavior. 45, 93–98 (2015).
17. Fernández-Alemán, J.L., Señor, I.C., Lozoya, P.Á.O., Toval, A.: Security and Privacy in Electronic Health Records: A Systematic Literature Review. Journal of Biomedical Informatics. 46, 541–562 (2013).
18. Ermakova, T., Fabian, B., Zarnekow, R.: Security and Privacy System Requirements for Adopting Cloud Computing in Healthcare Data Sharing Scenarios. In: Proceedings of the 19th Americas Conference on Information Systems. Chicago, IL, USA (2013).
19. Crossler, R.E., Bélanger, F.: The Mobile Privacy-Security Knowledge Gap Model: Understanding Behaviors. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 4071–4080. HI, USA (2017).
20. Flavián, C., Guinalíu, M.: Consumer Trust, Perceived Security and Privacy Policy: Three Basic Elements of Loyalty to a Website. Industrial Management & Data Systems. 106, 601–620 (2006).
21. Norberg P.A., Horne, D.R., Horne, D.A.: The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. The Journal of Consumer Affairs. 41, 100–126 (2007).
22. Crabtree, A., Tolmie, P., Knight, W.: Repacking 'Privacy' for a Networked World. Computer Supported Cooperative Work: CSCW: An International Journal. 26, 453–488 (2017).
23. Grobauer, B., Walloschek, T., Stöcker, E.: Understanding Cloud Computing Vulnerabilities. IEEE Security and Privacy. 9, 50–57 (2011).
24. Adelmeyer, M., Walterbusch, M., Biermanski, P., Teuteberg, F.: Trust Transitivity and Trust Propagation in Cloud Computing Ecosystems. In: Proceedings of the 26th European Conference on Information Systems. Portsmouth, UK (2018).
25. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An Integrative Model of Organizational Trust. Academy of Management Review. 20, 709–734 (1995).
26. Söllner, M., Hoffmann, A., Leimeister, J.M.: Why Different Trust Relationships Matter for Information Systems Users. European Journal of Information Systems. 25, 274–287 (2016).

27. Walterbusch, M., Martens, B., Teuteberg, F.: Exploring Trust in Cloud Computing: A Multi-Method Approach. In: Proceedings of the 21st European Conference on Information Systems (ECIS 2013). Utrecht, Netherlands (2013).
28. Mitchell, V.-W.: Consumer Perceived Risk: Conceptualisations and Models. European Journal of Marketing. 33, 163–195 (1999).
29. Hsin Chang, H., Wen Chen, S.: The Impact of Online Store Environment Cues on Purchase Intention: Trust and Perceived Risk as a Mediator. Online Information Review. 32, 818–841 (2008).
30. Acquisti, A., Brandimarte, L., Lowenstein, G.: Privacy and Human Behavior in the Age of Information. Science. 347, 509–514 (2015).
31. Lee, Y.E., Benbasat, I.: The Influence of Trade-off Difficulty Caused by Preference Elicitation Methods on User Acceptance of Recommendation Agents across Loss and Gain Conditions. Information Systems Research. 22, 867–884 (2011).
32. Walter, N., Öksüz, A., Walterbusch, M., Teuteberg, F., Becker, J.: "May I help You?" Increasing Trust in Cloud Computing Providers through Social Presence and the Reduction of Information Overload. In: Proceedings of the 35th International Conference on Information Systems. Auckland, NZ (2014).
33. Mathieson, K.: Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. Information Systems Research. 2, 173–191 (1991).
34. Pavlou, P.A., Fygenson, M.: Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly. 30, 115–143 (2006).
35. Aguinis, H., Bradley, K.J.: Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. Organizational Research Methods. 17, 351–371 (2014).
36. Shadish, W.R., Cook, T.D., Campbell, D.T.: Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston, NY (2002).
37. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. Information Systems Research. 13, 334–359 (2002).
38. Podsakoff, P.M., Organ, D.W.: Self-Reports in Organizational Research: Problems and Prospects. Journal of Management. 12, 531–544 (1986).
39. Burns, R.P., Burns, R.: Business Research Methods and Statistics using SPSS. Sage, London (2008).
40. Jarvis, C.B., MacKenzie, S.B., Podsakoff, P.M.: A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. Journal of Consumer Research. 30, 199–218 (2003).
41. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis. Pearson, Harlow, GB (2014).
42. Nunnally, J.C., Bernstein, I.H.: Psychometric Theory. McGraw-Hill, New York, NY, USA (1994).
43. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research. 18, 39–50 (1981).
44. Ringle, C.M., Wende, S., Becker, J.: SmartPLS 3, http://www.smartpls.com. (2015).
45. Hornburg, C., Giering, A.: Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. Marketing ZfP. 18, 5–24 (1996).
46. Hu, L., Bentler, P.M.: Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 6, 1–55 (1999).

Patientenintegration durch Pfadsysteme

Martin Benedict[1], Hannes Schlieter[1], Martin Burwitz[1], Tim Scheplitz[1], Marcel Susky[1], Peggy Richter[1], und Tjalf Ziemssen[2]

[1] Technische Universität Dresden, Fakultät Wirtschaftswissenschaften,
Professur für Wirtschaftsinformatik, insbes. Systementwicklung, Dresden, Deutschland
{vorname.nachname,peggy.richter2}@tu-dresden.de
[2] Universitätsklinikum Dresden, Zentrum für klinische Neurowissenschaften,
Multiple Sklerose Zentrum, Dresden, Deutschland
tjalf.ziemssen@uniklinikum-dresden.de

Abstract. Die aktive Integration eines Patienten in seinen Behandlungsverlauf besitzt große Potentiale zur Verbesserung der Behandlungsqualität. Als Ansatz zur besseren Patientenintegration werden häufig Patientenportale genannt. Die bestehenden Lösungen fokussieren jedoch vornehmlich auf die Zugänglichkeit elektronischer Patientenakten aber nicht auf die Integration von einrichtungsübergreifenden Informationen zum Behandlungsverlauf, auch Patientenpfad genannt. Der Beitrag schlägt daher eine Referenzarchitektur zur Integration von Pfadsystemen in Patientenportale vor. Ihre Anwendbarkeit wird anhand der Gestaltung eines Patientenportals für die Multiple Sklerose Behandlung demonstriert.

Keywords: Patientenpfade, Patientenportale, Referenzarchitektur, Integration

1 Einleitung

In der aktuellen Diskussion um die Verbesserung der Kontinuität der medizinischen Versorgung nehmen die personalisierte Medizin und die Förderung der Selbstbestimmung des Patienten (Patient Engagement) einen festen Platz ein. Dabei spielt auch die informatorische Integration des Patienten in den Behandlungsprozess eine zentrale Rolle. Über diese Form der Integration werden die Optimierung der Therapieadhärenz, positive Verhaltensänderungen beim Patienten sowie die Verbesserung von Prozess- und Ergebnisqualität erwartet. Sie bildet die Grundlage für die Einflussnahme des Patienten auf seinen Behandlungsprozess. In der Konsequenz soll der patientenindividuelle Outcome verbessert werden [1, 2]. Dies setzt jedoch voraus, dass der Patient seine Behandlung adäquat verfolgen kann, intendierte Behandlungspläne inklusive Falldokumentation zur Verfügung stehen und der Patient in der Lage ist, seinen Gesundheitsstatus nachzuvollziehen. Personal Health Records, Gesundheitsapps und Patientenportale [3, 4] ermöglichen grundsätzlich die informatorische Integration [5]. Dahingegen steht die Integration des Patienten in den Behandlungsprozess, d. h. die Bereitstellung und Mitwirkung des Patienten an einrichtungsübergreifenden Patientenpfaden und somit die Förderung des Selbstmanagements, noch am Anfang [6].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Der vorliegende Beitrag geht der Frage nach, wie einrichtungsübergreifende Patientenpfade im Kontext von Patientenportalen implementiert werden können. Dabei werden Patientenportale als Anwendungssysteme betrachtet, die die Integration des Patienten in die Informationssystemlandschaft mehrerer Leistungserbringer zulassen. Patientenportale können in verschiedenen Realisierungsformen auftreten (z. B. Web-Präsenz oder Apps). Die Schwerpunkte der Untersuchung sind die Erarbeitung eines Anforderungskatalogs und die Ableitung eines Referenzarchitekturvorschlags zur standardbasierten Implementierung einer pfadunterstützenden Anwendungssystemlandschaft (PAS). Im Zuge des Beitrags wird die Anwendung des Referenzarchitekturvorschlags anhand eines Patientenportals für die Versorgung von Multiple Sklerose (MS) Patienten demonstriert. Forschungsmethodisch ordnet sich die Arbeit in die gestaltungsorientierte Forschung ein [7] und zielt auf die Erarbeitung eines technischen Artefakts ab. Um dem Forschungsanspruch einer gestaltungsorientierten Theoriebildung [8] gerecht zu werden, wird die Referenzarchitektur als gestaltungsleitendes Artefakt betrachtet, das in verschiedenen Anwendungssystemlandschaften instanziiert werden kann. Etablierte Standards sowie eine rollenorientierte Beschreibung der Anwendungskomponenten ohne Konkretisierung der Systeminstanzen bilden dabei die Grundlage. Durch die rollenorientierte Beschreibung wird sowohl ein Abgleich mit existierenden Systeminstanzen (z. B. bestehende Anwendungssysteme) als auch die Entwicklung neuer Komponenten ermöglicht.

Die Struktur der Arbeit folgt dem von Peffers et al. postulierten Vorgehen für die Design Science Forschung [9]. Abschnitt 1 motiviert das Problemfeld und stellt die Problemklasse dar. In Kapitel 2 wird auf die Rolle von Patientenpfaden im Zusammenhang mit Patient Engagement und Patient Empowerment eingegangen (theoretische Basis). Aufbauend auf dem Gestaltungsziel werden in Abschnitt 3 die zentralen Anforderungen erarbeitet (Ziele der Lösung) und diese in Abschnitt 4 in einen Referenzarchitekturvorschlag als Design-Artefakt überführt. In Abschnitt 5 wird die vorgeschlagene Referenzarchitektur für einen Anwendungsfall instanziiert (Demonstration) [9].

2 Grundlagen

In den letzten Jahren findet ein verstärkter Diskurs zu den Konzepten, Maßnahmen und Auswirkungen einer stärkeren Einbindung des Patienten in die eigene Gesundheitsversorgung statt [10]. Zwei dabei adressierte Konzepte sind Patient Empowerment und Patient Engagement. Ersteres beschreibt die Ermächtigung des Patienten, mehr Verantwortung für seine gesundheitsrelevanten Entscheidungen zu übernehmen [11, 12]. Hingegen zielt Patient Engagement auf das Verhalten eines Patienten ab [1, 13]. Obgleich diskutabler Evidenz herrscht Einigkeit über die Vorteilhaftigkeit des Patient Engagement für die Verbesserung der Versorgungsqualität und -organisation [14], insbesondere in der Versorgung chronischer Erkrankungen [15]. Die Information, Edukation und Partizipation des Patienten werden dabei als essentielle Bestandteile hervorgehoben [11, 16]. Das aktive Gesundheitsverhalten eines Patienten im Sinne des Patient Engagement äußert sich somit in seiner informatorischen Teilhabe, welche wiederum eine optimale Kommunikation zwischen Versorgungsteam und Patient voraussetzt [13].

Patientenpfade stellen ein wesentliches Instrumentarium zur Verbesserung sowohl des Patient Empowerment als auch des Patient Engagement dar. Sie gehen auf das Konzept des Behandlungspfads (kurz: Pfad) zurück. Ein Pfad wird als standardbasierter Soll-Behandlungsablauf verstanden, der das multidisziplinäre Setting, die lokalen Gegebenheiten und den aktuellen Stand der Evidenzforschung vereint. Im Vordergrund steht die Vorausplanung konkreter Handlungsschritte, die an zeitliche bzw. definierte Zustandsänderungen geknüpft sind [17, 18]. Patientenpfade heben darüber hinaus den Patientenbezug hervor. Sie dienen insbesondere der nahtlosen Versorgung chronisch kranker Patienten über verschiedene Gesundheitssektoren hinweg, der Patienteninformation und -aufklärung sowie der Planung des patientenindividuellen Behandlungsprozesses unter Einbeziehung individueller Patientenbedürfnisse [19, 20].

Die Realisierung von Patienten Engagement wird mit Blick auf europäische Vorreiternationen und ihren Patientenportalen wie z. B. Dänemarks sundhed.dk oder Estlands e-tervis.ee deutlich. Im Mittelpunkt stehen hierbei die digitale Bereitstellung von medizinischen Dokumenten, Informationen über empfangene Behandlungen sowie deren Freigabe an Bevollmächtigte und weitere Dienste, wie Terminvereinbarungen, Medikamentenverschreibungen oder Kommunikationsangebote. Für Deutschland wird aktuell neben regionalen Initiativen das von vielen Krankenkassen unterstützte Patientenportal „Vivy“[1] diskutiert. Die Integration des Betroffenen als Mitgestalter seines individuellen Patientenpfades ist dabei jedoch bislang kaum berücksichtigt.

3 Anforderungsanalyse

Gleichwohl die Studienlage [18] und das politische Einfordern von Pfadsystemen (z. B. [21]) grundsätzlich für die Anwendung, und damit für die technische Unterstützung sprechen [22, 23], sind die Akzeptanz und Durchsetzung in der Praxis mangelhaft, woraus sich die folgenden Anforderungsbereiche ergeben.

3.1 Nutzeranforderungen für Pfadsysteme

Zur angemessenen Abbildung der Komplexität von Behandlungsprozessen sollten sich Werkzeuge zur Pfaderstellung und -pflege am fachlichen Kontext orientieren [24, 25] und alle Lebensphasen von Patientenpfaden unterstützen [26] (s. Tab. 1).

Tabelle 1. Behandlungsplanung

Nr.	*Beschreibung*
A1	**Leistungserbringer:** Zur Behandlungsplanung sollten typische Pfadvarianten und zugehörige Entscheidungsgrundlagen festgelegt werden können. Vorzugsweise sollte die Gestaltung und Editierung modellbasiert in einem integrierten Tooling erfolgen. **Patient:** Therapiemöglichkeiten und korrespondierende Pfadvarianten sollten für den Patienten erkennbar sein. Die möglichen Pfade sollten für den Patienten kognitiv erfassbar sein.

[1] https://www.vivy.com/

Nr.	Beschreibung
A2	**Leistungserbringer:** Für die Instanziierung eines Pfades sollten Ein- und Ausschlusskriterien hinterlegt werden. Der Patientenpfad sollte durch den Leistungserbringer individuell konfiguriert werden können. **Patient:** Der Patient sollte bei der Instanziierung des Pfades erkennen können, warum eine konkrete Pfadkonfiguration festgelegt wurde und hinsichtlich welcher Kriterien die Behandlung gestaltet wird.

Im ärztlichen Alltag ist es häufig notwendig, auf den Fallkontext und den Patientenwunsch zur Laufzeit eingehen zu können [27] (s. Tab. 2). Dadurch wird die Akzeptanz der Pfadnutzung gefördert.

Tabelle 2. Behandlungsorganisation

Nr.	Beschreibung
A3	**Leistungserbringer:** Durch den Leistungserbringer sollte jederzeit im Behandlungsverlauf eine ad-hoc Modifikation des Pfades möglich sein. Jede Modifikation sollte begründet werden, sodass es für den Fallkontext dokumentiert bleibt. **Patient:** Änderungen im Behandlungsverlauf sollten dem Patienten kommuniziert und die Gründe in nachvollziehbarer Weise dargestellt werden. Der Patient sollte in der Lage sein, Modifikationswünsche zu kommunizieren.
A4	**Leistungserbringer:** Die Nutzung von Pfaden sollte sich in ein dauerhaftes Prozessmanagement einbetten. D. h., die beteiligten Akteure begutachten und diskutieren in regelmäßigen Zyklen, inwieweit die Pfadvorlagen überarbeitet werden müssen. **Patient:** Änderungen von Pfadvorlagen sollten für den Patienten vor dem Hintergrund der eigenen Pfadinstanz transparent gemacht werden. Unterschiede zu Standardpfaden und nachträgliche Anpassungen sollten kommuniziert werden.

Mit der Nutzung von Behandlungspfaden gehen meist auch Wirtschaftlichkeitsüberlegungen einher [28]. Zudem wird sich eine höhere Sicherheit der Einhaltung von Prozessplänen erhofft. Diese Standardisierung in Form von Behandlungspfaden zieht jedoch gleichzeitig die Kritik des mangelnden Patientenbezugs (hinsichtlich Adäquanz und Zugänglichkeit) nach sich (s. Tab. 3) [29].

Tabelle 3. Behandlungsmonitoring

Nr.	Beschreibung
A5	**Leistungserbringer:** Der Leistungserbringer sollte patientenverständliche Informationen im Pfad hinterlegen können, welche der Edukation konkreter Patientenkohorten dienen. **Patient:** Der Patient sollte Zugang zu seinem Pfad haben und sich mittels verständlichen Statusinformationen und Zusatzmaterialien über den Fortgang seiner Behandlung informieren können.
A6	**Leistungserbringer:** Pfade sollten individualisiert und bei Bedarf um patientenindividuelle Informationen und Funktionen erweitert werden können. **Patient:** Pfadinformationen sollten mit korrespondierenden Assistenzfunktionen verknüpft sein und dem Patienten bedarfsgerecht und personalisiert zur Verfügung gestellt werden.

Nr.	Beschreibung
A7	**Leistungserbringer:** Informationen über den Ist-Zustand der Behandlung sollten durch den Leistungserbringer im Pfad hinterlegt und im Zuge eines prozessorientierten Qualitätsmanagements überprüft werden können. **Patient:** Der Patient sollte den aktuellen Zustand seiner eigenen Pfadinstanz vor dem Hintergrund der ursprünglich intendierten Behandlungsplanung analysieren und Abweichungen nachvollziehen können.

3.2 Systemanforderungen für Pfadsysteme

Auf Grundlage der Nutzeranforderungen A1-A7 sind in Tabelle 4 funktionale Anforderungen herausgearbeitet. Die funktionalen Anforderungen repräsentierten einen Basissatz, der in Abhängigkeit vom spezifischen medizinischen Einsatzszenario hinsichtlich Funktionsumfang und Integrationstiefe der Pfadanwendung variieren kann. Dieser Basissatz bestimmt den Rahmen für die zu gestaltende PAS. Die PAS ist dabei nicht als einzelnes Softwareprodukt zu verstehen, sondern als System von Anwendungssystemen, welches sich im Kern aus dem beteiligten klinischen Anwendungssystem (KAS) [30], dem Patientenportal mit seinen Modulen sowie externen Infrastrukturen, Systemen und Diensten konstituiert [31].

Tabelle 4. Funktionale (F) Anforderungen zur Umsetzung des Pfadsystems

Nr.	Beschreibung
F1	Die Aufgaben- und Terminsteuerung sollte in der PAS aktiv erfolgen. Termine sowie Aufgaben sollten aktiv in die beteiligten Systeme (KAS, Patientenportal etc.) kommuniziert werden können.
F2	Die PAS sollte in der Lage sein, beteiligte Anwendungen darüber zu informieren, wenn sich der Status eines Patienten, eines Fallverlaufs, einer Behandlung oder konkreter Werte ändert.
F3	Pfadvorlagen sollten mittels eines geeigneten Toolings (z. B. Modellierungswerkzeug) komfortabel erstellt und verwaltet werden können sowie den beteiligten Systemen zur Verfügung stehen.
F4	Die Pfadvorlagen sollten in die Falldokumentation der PAS integriert sein und patientenspezifisch instanziiert werden können.
F5	Patientenpfade sollten für berechtigte Akteure im Fallkontext des Patienten einsehbar und editierbar sein. Die Modifikation des geplanten patientenindividuellen Pfades sollte jederzeit möglich sein, aber auch die Möglichkeit zur Begründung von Modifikationen ist unerlässlich.
F6	Auswertungs- und Monitoring-Funktionalitäten sollten zentralisiert für jeden Stakeholder erfolgen können (Management-Cockpit). Alternativ sollten bestehende Process Mining Tools (z. B. ProM, RapideMiner, BAB) integrierbar sein.
F7	Pfadinstanzen sollten sich kontinuierlich der Dokumentationslage der beteiligten KAS und anderer Dokumentationssysteme anpassen.

4 Referenzarchitektur

Auf Basis der erhobenen Nutzer- und Systemanforderungen werden im Folgenden die Komponenten und Module im Sinne von Architekturbausteinen erarbeitet (Abs. 4.1) und eine Referenzarchitektur entwickelt (Abs. 4.2), die am Beispiel der Implementierung eines Patientenportals für die MS-Versorgung (Abs. 5) illustriert wird.

4.1 Module mit Pfadbezug in Patientenportalen

Der Funktionsumfang von Patientenportalen spiegelt die krankheitsspezifischen informatorischen Bedürfnisse eines Patienten wider [3]. Die Konfiguration bzw. Instanziierung eines Patientenportals können somit auch hinsichtlich Auswahl und Umsetzung der Funktionsumfänge variieren. Die in Tabelle 5 vorgeschlagenen Module beschreiben die Funktionalitäten, die für die Umsetzung von Patientenpfaden in Patientenportalen notwendig sind. Die Menge ist jedoch nicht als abgeschlossene Menge zu betrachten, da der Versorgungskontext weitere Module erforderlich machen kann.

Tabelle 5. Beschreibung der Patientenportalmodule

Modul	*Beschreibung*	*Referenz*
	Infrastruktur-/Kernelmodule	
Kernel	Dieses Modul implementiert die grundsätzlichen Funktionalitäten und Konfigurationen (Authentifizierung, Identitäts-, Demografisches Daten- und Softwareschnittstellen-Management sowie Metadaten-Konfiguration).	[32]
	Spezifische Pfadmodule	
Care Plan Module	Dieses Modul implementiert solche Pfadfunktionalitäten, die durch Patient Engagement beeinflussbar sind.	F1-F7
Self-Tracking Module	Dieses Modul implementiert das Monitoring des Gesundheitsstatus.	F6
	Pfadassoziierte Module (siehe Funktionsdomänen in [3])	
Encounter Module	Dieses Modul implementiert Funktionalitäten zum Management von Terminen und anderen ärztlichen sowie pflegerischen Interventionen.	F1, F2
Documentation Module	Dieses Modul implementiert die Zugriffsfunktionalitäten auf die medizinische Akte und andere medizinische Dokumentation sowie die Bereitstellung von Dokumentation durch den Patienten.	F2, F4
Assessment Module	Dieses Modul implementiert die Screening- und Fragebogen-Funktionalitäten.	F1, F7
Education Module	Dieses Modul implementiert die Funktionalitäten zur patientenindividuellen Therapieunterstützung (Übungsanleitungen, Erklärungsmaterialen etc.).	F5
Medication Module	Dieses Modul implementiert die medikationsbezogenen Verwaltungs- und Monitoring-Funktionalitäten.	F7

Grundlegend wird eine Unterteilung in Kernelmodule, spezifische Pfadmodule und pfadassoziierte Module[2] vorgenommen. Die *Kernelmodule*, auch Kernel genannt [32], beinhalten die Basiskonfiguration sowie die zentralen Funktionalitäten und Schnittstellen, von denen die Funktionalitäten der spezifischen Pfad- und pfadassoziierten Module abhängig sind (z. B. Rechteverwaltung). Die *spezifischen Pfadmodule* verantworten

[2] **Komponenten:** eigenständige Einheiten in einer Architektur,
Module: logische Teileinheiten von Komponenten

Funktionalitäten zur Abbildung und Ausführung konkreter Pfadinstanzen. *Pfadassoziierte Module* hingegen bilden fachliche Funktionalitäten mit Bezug zum Pfad ab. So kann sich bspw. ein Patient im pfadassoziierten Modul „Documentation“ alle Dokumente anzeigen lassen, welche mit dem Pfadschritt einer Anamnese verknüpft sind. Ebenso sind Termine oder verknüpfte Medikationen mit den Modulen „Encounter“ bzw. „Medication“ abrufbar.

Die Module können Schnittstellen zu externen Geräten (z. B. Blutdruckmessgeräte) und Anwendungen (z. B. Gesundheits-Apps) vorsehen. Da externe Systeme in verschiedenen Kontexten zum Einsatz kommen können, dienen sie als fallabhängige Integrationspunkte. Die Referenzarchitektur verweist für die Schnittstellen des Patientenportals explizit auf die Ressourcen des Fast Healthcare Interoperability Resources-(FHIR)-Standards[3], um Orientierung zu geben, welche Informationstypen das jeweilige Modul bereitstellen sollte. Grundsätzlich müssen Schnittstellen zu externen Anwendungen an das jeweilige Krankheitsbild und den Einsatzkontext angepasst werden, sodass auch andere Standards oder Formate zum Einsatz kommen können.

4.2 Referenzarchitektur zur Integration von Patientenportalen in PAS

Abbildung 1 stellt die Referenzarchitektur dar, in der die Rollen[4] einzelner Komponenten sowie die Schnittstellen zwischen diesen zusammengefasst werden. Dabei wird der Architekturvorschlag von Schlieter et al. [31] aufgegriffen und um die Kommunikationsbeziehungen zwischen Back-End-Diensten und Patientenportal erweitert. Zudem werden die Ausgestaltung der dort benannten Patient-Frontends konkretisiert und die Integration mit den vorgeschlagenen Pfadsystemen aufgezeigt. Die dargestellte Referenzarchitektur geht von einer einrichtungsübergreifenden Realisierung von Patientenpfaden und Falldokumentation aus. Dieses Implementierungsparadigma findet sich häufig bei der Umsetzung einrichtungsübergreifender Informationssysteme zum Patientendatenaustausch (z. B. [5]).

Entsprechend Abbildung 1 verantwortet das *„Patient-oriented Pathway Repository“ (PPR)* die Persistierung der Pfadinstanzen. Die Pfadvorlagen werden im *Template Repository (TR)* hinterlegt. Die professionelle Steuerung, d. h. die Verwendung der Pfade durch den Leistungserbringer, erfolgt durch den *Professional Pathway-Service (PrPS)*. Diese Komponente fasst dabei die in [31] erwähnten Komponenten zur Pfadinstanzierung, -modifikation sowie die Komponenten zur Datenaggregation und Pfadanalyse zusammen. Zudem setzt sie aktive Pfadveränderungen und Automatismen bei der Pfadausführung um und ermöglicht dem Leistungserbringer die Beeinflussung der Pfadausführung.

Die Komponente *Patient Pathway Service (PaPS)* dient der Ausführung von Pfaden aus der Sicht des Patienten. Diese Komponente setzt die pfadbezogenen Funktionen für

[3] http://hl7.org/fhir/

[4] Die in der Referenzarchitektur benannten Komponenten werden als Rollen betrachtet, die ein Anwendungssystem einnehmen kann. Beispielsweise, können die klinischen und professionellen Anwendungssysteme gleichzeitig als Komponenten in anderen Informationsinfrastrukturen, zum Beispiel einer elektronischen Akteninfrastruktur, enthalten sein.

den Patienten um. Sie ermöglicht bspw. den lesenden Zugang von Pfadinformationen. Gleichsam stellt sie Funktionalitäten zur Verfügung, die eine direkte Beteiligung des Patienten am Behandlungsprozess durch Modifikation von Pfadinstanzen ermöglichen. So können Fragebögen im Pfad hinterlegt werden, die in Abhängigkeit vom Behandlungsfortschritt durch den Patienten ausgefüllt werden müssen.

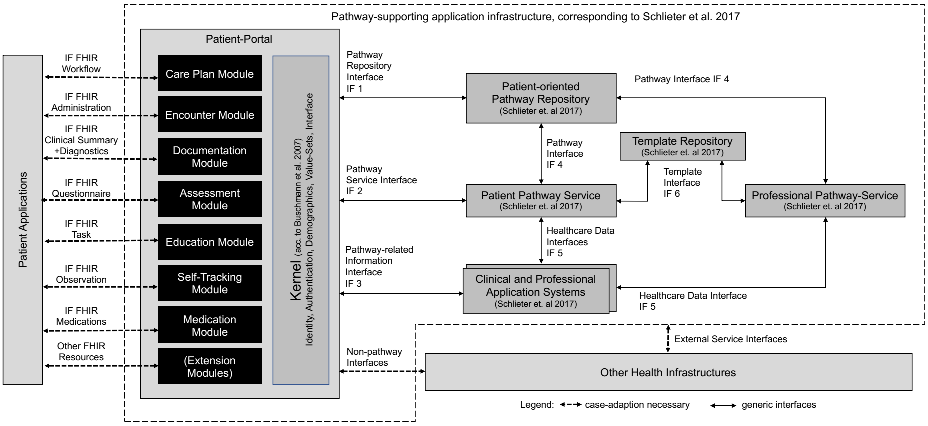


Abbildung 1. Referenzarchitektur zur Einbindung von Patientenportalen in PAS

Die Komponente *Clincial and Professional Application Systems (CAS)* repräsentiert die bestehenden klinischen Dokumentations- und Managementsysteme (KAS, Laborinformationssysteme, Arztpraxisinformationssysteme etc.). Diese Systeme sind Bestandteil der PAS, wenn sie Informationen und Funktionen bereitstellen, die für die Planung, Ausführung oder das Monitoring des Pfades Relevanz haben.

Zwischen den dargestellten Komponenten sind verschiedene bidirektionale Schnittstellentypen (siehe Abb. 1) definiert. Tabelle 6 stellt diese, die Informationen, sowie die Zugriffsarten dar. Die Zugriffsart „aktiv“ repräsentiert die Bereitstellung auf Anfrage. "Reaktiv" repräsentiert die unaufgeforderte Bereitstellung und automatisierte Verarbeitung als Reaktion auf ein auslösendes Ereignis.

Tabelle 6. Schnittstellentypen (SST) der Referenzarchitektur

SST	*Beschreibung*	*Information und Zugriff*
IF1 – Pathway Repository Interface	Die Schnittstelle dient dem direkten Zugriff auf die Rohdaten der Pfadinstanzen. Es handelt sich hierbei um den in [31] genannten Zugriff auf das Repository auf Basis des IHE XDS.b-Standards.	Vollständige Pfadinstanzen, aktiver Zugriff
IF2 – Pathway Service Interface	Die Schnittstelle setzt den Austausch von Pfadeinzelinformationen um. Das umfasst Schnittstellen zur Anpassung und Konsumierung von Pfadinformationen. Die Schnittstelle fungiert auch als Service-Schnittstelle zum Auslösen von Funktionen der Patienten-Pfadkomponente.	Statusänderungen, Pfadinformationen, aktiver und reaktiver Zugriff

SST	*Beschreibung*	*Information und Zugriff*
IF3 – Pathway-related Information Interface	Die Schnittstelle dient dem Austausch von pfadrelevanten Informationen. Das können Informationen sein, die mit der direkten Ausführung des Pfades assoziiert sind oder die Auswirkungen auf den Pfadverlauf haben.	Pfadbeeinflussende und pfadresultierende Fachinformationen, aktiver und reaktiver Zugriff
IF4 – Pathway Interface	Diese Schnittstelle repräsentiert den Zugriff der aktiven Pfad-Service-Komponenten auf die Pfadinstanzen.	Pfadinstanzen, aktiver Zugriff
IF5 – Healthcare Data Interfaces	Die Schnittstelle dient dem Austausch von Daten zur Beeinflussung von Pfadinstanzen bzw. zur Umsetzung von Pfadverläufen. Über diese Schnittstelle werden bspw. Daten kommuniziert, die Entscheidungen in einem Pfad beeinflussen.	Informationen, die den Pfadverlauf beeinflussen, aktiver und reaktiver Zugriff
IF6 – Template Interface	Das Template Interface dient der Bereitstellung und Modifikation von Pfadvorlagen.	Pfadvorlagen, Anpassung von Pfadvorlagen, aktiver Zugriff
External Service Interfaces	Die Pfad-Services können bei der Ausführung von weiteren Infrastrukturen (z. B. Anwendungssystemlandschaften für den Labordatenaustausch) abhängig sein.	Variabel, aktiver und reaktiver Zugriff
Non-pathway Interfaces	Pfade und korrespondierende Module des Patientenportals können auf konkrete Informationen und Funktionen verweisen. Diese Informationen beeinflussen jedoch den Pfadverlauf nicht oder führen nicht zu Änderungen im Pfad. Schnittstellen zur Kommunikation solcher Informationen werden daher als Non-pathway Interfaces klassifiziert.	Variabel vorrangig aktiver Zugriff in Bezug auf die Pfadausführung

5 Demonstration – Fallbeispiel Multiple Sklerose

5.1 Fallkontext – Integrierte Multiple Sklerose Versorgung

Die MS ist eine chronisch entzündliche, degenerative Erkrankung des zentralen Nervensystems [33]. Als „Krankheit der 1000 Gesichter“ ist sie gekennzeichnet durch ein heterogenes, stark beeinträchtigendes Beschwerdebild [34]. Aus dem jungen Eintrittsalter resultieren auch Eigenschaften der Patienten, welche insbesondere im Kontext moderner, partizipativer Versorgungsansätze fördernd wirken. So wird jüngeren Betroffenen eine IT-Affinität und höhere Kompetenz in der Erschließung sowie Nutzung elektronischer Gesundheitsangebote zur Förderung der eigenen Patientenkompetenz zugesprochen [35, 36].

Im Zuge des Projektes „Integriertes Betreuungsportal Multiple Sklerose“ (IBMS) wird ein Patientenportal für MS-Patienten entwickelt, das die Heterogenität und Individualität der Erkrankung berücksichtigen soll. Das Patientenportal soll sich in die bestehende Informationssystemlandschaft am Universitätsklinikum Dresden (UKD) einbetten und darüber hinaus andere Leistungserbringer der sächsische Gesundheitswirtschaft einschließen. Ein Teilziel des Projektes ist es, bei der Entwicklung des Patientenportals die bestehenden Systeme des UKD sowie der Region in Form einer Integration zu berücksichtigen. Zu den existierenden Systemen gehören das am UKD eingesetzt Fachanwendungssystem Multiple Sklerose Dokumentationssystem^{3D} (MSDS^{3D}) zur Dokumentation von MS-Behandlungen und die IHE-basierte CCS Telehealth Ostsachsen Plattform (THOS) zur Implementierung einrichtungsübergreifender Telemedizin-Lösungen. Die Patientenzugänglichkeit über ein Patientenportal ist ein Teilziel des Projektes. Ein Patient soll sowohl seinen Behandlungsverlauf, seine medizinische Dokumentation als auch seinen Erkrankungsstatus nachvollziehen können.

5.2 Applikation der Referenzarchitektur

Im Rahmen des Fallkontextes wird auf Basis der vorgestellten Referenzarchitektur ein Patientenportal für das beteiligte MS-Zentrum konzipiert und implementiert (vgl. Abb. 2). Dazu wurden die bestehenden Anwendungssysteme analysiert und mit den Rollen der Referenzarchitektur abgeglichen. Vorhandene Anwendungssysteme sind die XDS.b-basierte Fallakte und der Integrated Clinical Pathway (ICP)-Service, welche die Rollen *PPR* bzw. *PrPS, PaPS* und *TR* der Referenzarchitektur umsetzen, sowie das Klinisches Anwendungssystem MSDS^{3D} (s. Tab. 7), welches die Rolle des *CAS* realisiert. Vorgesehen ist außerdem, externe Praxisinformationssystem in der Rolle von *CAS* in die Lösung einzubinden. Der Patient Pathway-Manager wird im Zuge des Projektes entwickelt und in die Anwendungssystemlandschaft integriert.

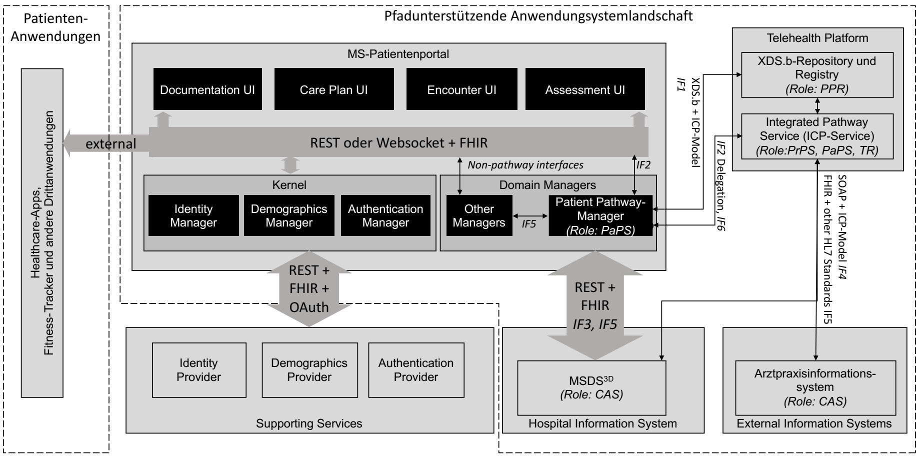


Abbildung 2. Anwendungssystemarchitektur für ein MS-Patientenportal

Die aktuelle Anwendungssystemarchitektur umfasst die Präsentationsschicht, welches dem User Interface (UI) entspricht und die fachlichen Module der Referenzarchitektur (s. Abb. 1, links) repräsentiert. Das Backend besteht aus einem erweiterbaren *Kernel*, der das Kernel-Modul der Referenzarchitektur (s. Abb. 1, links) repräsentiert, sowie aus verschiedenen *Domain-Managern*. Die Präsentationsschicht ist vom Backend über standardisierte REST-basierte FHIR- bzw. Websocket-Schnittstellen entkoppelt.

Dadurch fungiert das Backend gleichzeitig als Middleware, welche die Anbindung von Drittanwendungen (s. Abb. 2, links), z. B. das FHIR-basierte "Apple Health Records" [37], ermöglicht.

Mit Bezug zum Patienten verantwortet der *Kernel* die systemweit eindeutige Identifikation, die Akquise und Bereitstellung demografischer Daten sowie die Authentifizierung und Zugriffskontrolle am Patientenportal. Der *Kernel* ist so gestaltet, dass zur Wahrnehmung seiner Verantwortlichkeiten auf externe Dienste zurückgegriffen werden kann. Die *Domain-Manager* setzen die domänenspezifische Geschäftslogik um, wobei ein UI-Modul zumindest durch einen *Domain-Manager* verantwortet wird. Der *Patient Pathway-Manager* realisiert die Logik für das Care Plan Modul und nutzt die vorhandenen anderen *Domain-Manager*, welche den Zugang zu den jeweiligen Fachdaten aus den CAS gewährleisten.

Tabelle 7. Komponenten und deren Schnittstellen in der MS-Patientenportalarchitektur

Komponente *(Status der Verfügbarkeit) - Aufgabe;* ***Schnittstelle:*** *Typ: Datenmodell*	*Rolle*
Patient Pathway-Manager (zu implementieren) – Patientenseitige Ausführung von Pfaden **Schnittstellen:** *IF1:* Proprietäre Beschreibung (CPMod-Format) [24], FHIR Workflow-Ressourcen[5] (aktiv), *IF2:* ICP-Service im Sinne eines Business Delegates [32], SOAP-Calls proprietär (aktiv).	PaPS
XDS.b-Registry und Repository (existent) – Speicherung von Patientendaten und -pfaden **Schnittstellen:** *IF1, IF4:* CPMod für Patientenpfade, Repository-Schnittstellen (XDS.b-basierter Austausch, aktiv)	PPR
ICP-Service (existent) – Ausführung von Pfaden auf Seiten der Leistungserbringer **Schnittstellen:** *IF2:* FHIR Workflow-Ressourcen (aktiv und reaktiv), *IF4, IF6*: CPMod (aktiv) *IF5:* FHIR-Ressourcen: Observation (aktiv und reaktiv), Ereignisse (FHIR-Ressourcen: Events[6], reaktiv), Appointment und Task[7] (aktiv und reaktiv)	PrPS PaPS
MSDS^{3D} (existent) – Dokumentation von MS-Fällen **Schnittstellen:** *IF3*: FHIR-Communication-Ressourcen (reaktiv), Kommunikation von Patientendaten (Non-pathway Interfaces).	CAS
UI-Module Documentation, Care Plan, Encounter, Assessment (zu implementieren) – Umsetzung von Patientenfunktionen **Schnittstellen:** *intern und extern:* FHIR-Ressourcen gemäß Schnittstellen zu externen Anwendungen in Referenzarchitektur (siehe Abb. 1) auf Basis JSON + Websocket, alternativ REST	Module
Kernel-Module Identity Manager, Demographics Manager, Authentication Manager (zu implementieren) – Anbindung externer Dienste **Schnittstellen:** *Non-pathway Interfaces:* FHIR-Administration-Ressourcen, OAuth, Json Web Token (JWT)/JSON-Format	Kernel

Zur Komposition der einzelnen Anwendungssysteme in ihrer jeweiligen Rolle (s. Abb. 1) müssen Schnittstellen zur Anbindung an das Patientenportal im Sinne der Referenzarchitektur (s. Tab. 6) geschaffen werden. Die semantische Interoperabilität wird

[5] https://www.hl7.org/fhir/workflow.html
[6] https://www.hl7.org/fhir/event.html
[7] https://www.hl7.org/fhir/task.html, https://www.hl7.org/fhir/appointment.html

sichergestellt, indem auf die vorhandenen REST-basierten FHIR- und IHE-XDS.b-basierte Schnittstellen zurückgegriffen wird (s. Tab. 7). Die Grundlage dafür ist die Erstellung eines abgestimmten Kommunikationsdatenmodells, bei dem die kontextspezifischen Informationsbedarfe von Patienten Berücksichtigung finden. Weiterhin wird auf das Clinical-Pathway-Modeling-Format (CPmod) aus [24] zurückgegriffen.

Nach aktuellem Entwicklungsstand des Patientenportals navigiert der Patient über den Menüpunkt „Meine Behandlung“ in *Care Plan UI*, in dem er entlang eines Zeitstrahls sowohl auf historische Daten als auch auf geplante Aktivitäten zugreifen kann (s. Abb. 3). Die Informationen werden über den *Patient Pathway-Manager* bereitgestellt, der wiederum unter Nutzung des *Kernels* die Datenintegrität sowie die Authentizität des Nutzers sicherstellt. Der *Patient Pathway-Manager* kann sich die Informationen zur korrespondierenden Pfadinstanz (*IF1*) sowie zur Pfadvorlage des Patienten (*IF6*) über die THOS abrufen und so den Patienten informieren, welche Schritte im Standardpfad vorgesehen sind oder Abweichungsanalysen auslösen.

Auf die Falldokumentation im klinischen Dokumentationssystem kann ebenso mittels REST-Schnittstelle (*IF3*) zugegriffen werden. Somit lassen sich bspw. Laborergebnisse mit den Pfadschritten verknüpfen. Der Patient erhält damit die Informationen, die im aktuellen Pfadschritt relevant sind. Gleichzeitig können Anfragen, Ereignisse oder Hinweise an den Patienten gesendet werden. Für das Patientenportal wird im Projekt über diesen Mechanismus eine Fragebogenkomponente integriert, über die Informationen zum Zustand des Patienten abgefragt werden können.

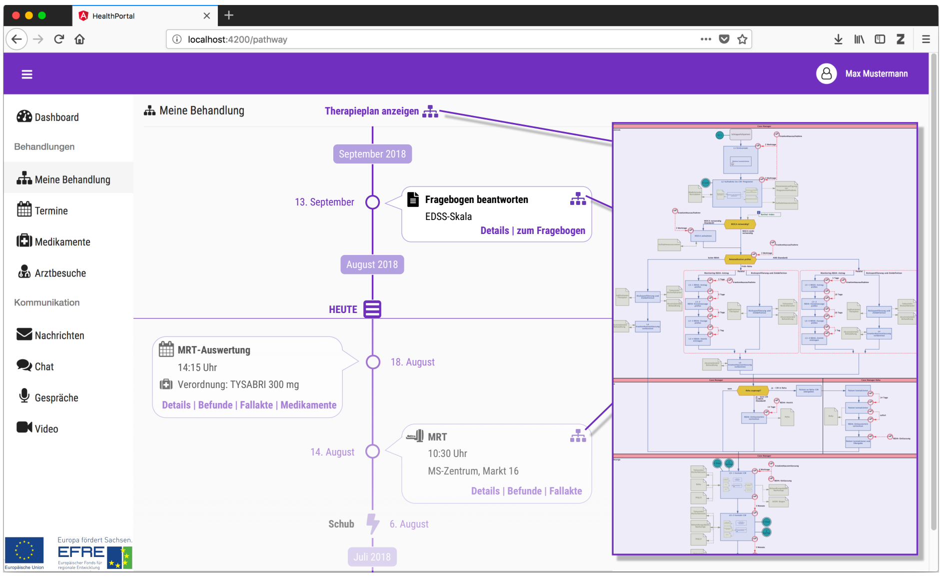


Abbildung 3. Screenshot des Patientenportals inklusive schematisch angedeutete Integration der Pfadvorlage

6 Diskussion und Fazit

Der vorliegende Beitrag widmete sich der Umsetzung von Patientenpfaden und speziell ihrer Integration in den Kontext eines Patientenportals. Dazu wurden auf Grundlage eines Anforderungskatalogs architektonische Überlegungen in eine Referenzarchitektur überführt, in welcher die typischen Systemkomponenten, deren Rollen sowie typische Kommunikationsbeziehungen spezifiziert sind. Auf Basis eines Proof-of-Concepts in der MS-Versorgung wurde sowohl die Anwendbarkeit als auch Instanziierung des generischen Architekturkonzepts aufgezeigt [38]. Bereits während der Entwicklung des Patientenportals konnten die notwendigen Funktionsbereiche eingegrenzt werden (Patientenbefragung mit n=210[8]). Die Evaluierung der Referenzarchitektur sowie der fallbezogenen Gesamtlösung aus Nutzersicht steht jedoch noch aus und wird Gegenstand der Folgearbeit sein.

Zudem muss konstatiert werden, dass der Beitrag auf die funktional-technologische Realisierung des Portals abhebt. In der Praxis zeigt sich jedoch, dass auch mit organisatorischen (Finanzierung, Einführungs-, Schulungsfragen sowie Support) und datenschutzrechtlichen Aspekten eine Vielzahl von Einführungshürden einhergehen, die im Beitrag ausgeklammert wurden. Hier kann anregt werden, den Konsens für standardisierte Verfahren zu befördern und den rechtlichen Rahmen an die digitalen Möglichkeiten anzupassen. Zentrale elektronische Akten, wie TK-safe[9] in Deutschland und die finnischen KANTA Services[10] auf europäischer Ebene, geben Einblicke wie Patientenintegration angegangen werden kann. Daraus ergeben sich Potentiale für die Umsetzung pfadgestützter Systeme, für die die aufgezeigte Referenzarchitektur als Entwurfshilfe dienen kann. Mit dem zusammengeführten Anforderungskatalog sowie den elaborierten Schnittstellentypen gibt die Arbeit eine praktische Handreiche, um den Transformationsprozess hin zu einer pfadzentrierten Patientenintegration zu gestalten.

Trotz der weitestgehend positiven Konnotation des Themas Patient Engagement existiert insbesondere um die Fragen, inwieweit bereits ohne ärztlichen Kontakt Befunde bereitgestellt oder Empfehlungen automatisiert getroffen werden können, ein ethischer Diskurs. Dieser wird in den nächsten Jahren verstärkt geführt werden müssen.

Danksagung

Das Universitätsklinikum Carl Gustav Carus Dresden, die Technische Universität Dresden und die CCS GmbH kooperieren im Rahmen des Projektes „Integriertes Betreuungsportal Multiple Sklerose". Die Durchführung des Projektes wird durch Mittel des Europäischen Fonds für Regionale Entwicklung (EFRE) und den Freistaat Sachsen ermöglicht. Der Beitrag ist dem gesamten helict-Team der TU Dresden gewidmet, welches in den vergangenen Jahren unermüdlich für die Überwindung tradierter Verhaltensreflexe und für die Entwicklung innovativer digitaler Lösungen gestritten hat.

[8] https://bit.ly/2PUj9p9
[9] https://www.tk.de/tk/digitale-gesundheit/
[10] https://www.kanta.fi/en/my-kanta-pages

Referenzen

1. Carman, K.L., Dardess, P., Maurer, M., Sofaer, S., Adams, K., Bechtel, C., Sweeney, J.: Patient And Family Engagement: A Framework For Understanding The Elements And Developing Interventions And Policies. Health Aff. (Millwood). 32, 223–231 (2013).
2. Freudenthaler, I.: Der zufriedene Patient: Qualitäts- und Praxismanagement für den Arzt. Springer-Verlag (2011).
3. Scheplitz, T., Benedict, M., Esswein, W.: Patientenkompetenz durch Online-Portale - Eine Funktionsanalyse. In: Proceedings MKWI 2018. , Lüneburg (2018).
4. Ammenwerth, E., Schnell-Inderst, P., Hoerbst, A.: Patient empowerment by electronic health records: first results of a systematic review on the benefit of patient portals. Stud. Health Technol. Inform. 165, 63–67 (2011).
5. Nina, L., Sabine, K., Maria, H., Peter, B., Nadia, D., Johan, E., Olof, J., Anja, P., Vivian, V., Casper, W.: My Care Pathways - Creating Open Innovation in Healthcare. Stud. Health Technol. Inform. 687–691 (2013).
6. Manuela Plößnig, Yildiray Kabak, Ilias Lamprinos, Alexander Pabst, Claudia Hildebrand, Sarah Mantwill: EMPOWER - Pathways for Supporting the Self-management of Diabetes Patients. Stud. Health Technol. Inform. 159–166 (2015).
7. Winter, R.: Design science research in Europe. Eur. J. Inf. Syst. 17, 470–475 (2008).
8. Kuechler, W., Vaishnavi, V.: A Framework for Theory Development in Design Science Research: Multiple Perspectives. J. Assoc. Inf. Syst. 13, (2012).
9. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 24, 45–77 (2007).
10. Risling, T., Martinez, J., Young, J., Thorp-Froslie, N.: Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. J. Med. Internet Res. 19, e329 (2017).
11. Gouthier, M., Tunder, R., h: Die Empowerment-Bewegung und ihre Auswirkungen auf das Gesundheitswesen. In: Hoefert, H.-W. (ed.) Wandel der Patientenrolle. Hogrefe, Göttingen [u.a.] (2011).
12. Kambhampati, S., Ashvetiya, T., Stone, N.J., Blumenthal, R.S., Martin, S.S.: Shared Decision-Making and Patient Empowerment in Preventive Cardiology. Curr. Cardiol. Rep. 18, 49 (2016).
13. Deering, M.J., Baur, C.: Patient portals can enable provider-patient collaboration and person-centered care. In: Grando, M.A., Rozenblum, R., and Bates, D.W. (eds.) Information technology for patient empowerment in healthcare. pp. 130–152 (2015).
14. Barello, S., Graffigna, G., Vegni, E.: Patient Engagement as an Emerging Challenge for Healthcare Services: Mapping the Literature. Nurs. Res. Pract. (2012).
15. Gensichen, J., Rosemann, T.: Das Chronic Care Modell: Elemente einer umfassenden Behandlung für Patienten mit chronischen Krankheiten DEGAM-Serie: Betreuung von Patienten mit chronischen Krankheiten. ZFA - Z. Für Allg. 83, 483–486 (2007).
16. Schaeffer, D., Haslbeck, J.: Bewältigung chronischer Krankheit. In: Richter, M. and Hurrelmann, K. (eds.) Soziologie von Gesundheit und Krankheit. pp. 243–256. Springer Fachmedien Wiesbaden, Wiesbaden (2016).
17. De Bleser, L., Depreitere, R., De Waele, K., Vanhaecht, K., Vlayen, J., Sermeus, W.: Defining pathways. J. Nurs. Manag. 14, 553–563 (2006).
18. Lawal, A.K., Rotter, T., Kinsman, L., Machotta, A., Ronellenfitsch, U., Scott, S.D., Goodridge, D., Plishka, C., Groot, G.: What is a clinical pathway? Refinement of an operational definition to identify clinical pathway studies for a Cochrane systematic review. BMC Med. 14, (2016).

19. Martin, H.M., Pedersen, L.H., Johansen, A.M.U.: Method description: user-led healthcare in patient pathways. Ugeskr. Laeger. 179, (2017).
20. Coughlan, J., Eatock, J., Eldabi, T.: Evaluating telemedicine: a focus on patient pathways. Int. J. Technol. Assess. Health Care. 22, 136–142 (2006).
21. Albreht, T., Kiasuwa, R., Van der Bulcke, M. eds: EUROPEAN guide on quality improvement in comprehensive cancer control. Ljubljana: National Institute of Public Health; Brussels: Scientific Institute of Public Health (2017).
22. Rotter, T.: Clinical Pathways in Hospitals: Evaluating effects and costs, http://repub.eur.nl/pub/39428/, (2013).
23. Lenz, R., Reichert, M.: IT support for healthcare processes – premises, challenges, perspectives. Data Knowl. Eng. 61, 39–58 (2007).
24. Burwitz, M., Schlieter, H., Esswein, W.: Modeling Clinical Pathways-Design and Application of a Domain-Specific Modeling Language. In: Alt, A. and Franczyk, B. (eds.) Tagungsband der 11. Internationalen Tagung Wirtschaftsinformatik. , Leipzig (2013).
25. Heß, M., Kaczmarek, M., Frank, U., Podleska, L., Täger, G.: A Domain-specific Modelling Language for Clinical Pathways in the Realm of Multi-perspective Hospital Modelling. In: ECIS 2015 (2015).
26. Juhrisch, M., Schlieter, H., Dietz, G.: Information systems engineering in healthcare - an evaluation of the state of the art of operational process design. Int. J. Organ. Des. Eng. 2, 420–444 (2012).
27. Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S.: Evidence based medicine: what it is and what it isn't. BMJ. 312, 71–72 (1996).
28. Fischer, B., Küppers-Simon, M.: Design des Pfades. In: Eckardt, J. and Sens, B. (eds.) Praxishandbuch Integrierte Behandlungspfade. pp. 81–122. Economica Verlag (2006).
29. Schriek, M., Turetken, O., Kaymak, U.: A Maturity Model for Care Pathways. In: Proceeding ECIS2016. , Istanbul, Turkey (2016).
30. Haas, P.: Medizinische Informationssysteme und Elektronische Krankenakten. Springer Berlin Heidelberg (2005).
31. Schlieter, H., Benedict, M., Gand, K., Burwitz, M.: Towards Adaptive Pathways: Reference Architecture for Personalized Dynamic Pathways. Presented at the CBI July (2017).
32. Buschmann, F.: Pattern-oriented software architecture: a Pattern Language for Distributed Computing. Wiley, Chichester [England]; New York (1996).
33. Kip, M., Zimmermann, A., Bleß, H.: Epidemiologie der Multiplen Sklerose. In: Kip, M., Schönfelder, T., and Bleß, H. (eds.) Weißbuch Multiple Sklerose. pp. 13–22. Springer Berlin Heidelberg and Imprint: Springer, Berlin, Heidelberg (2016).
34. Yaldizli, Ö., Kappos, L.: Klinische Grundlagen der Multiplen Sklerose. In: Egli, S. (ed.) Multiple Sklerose verstehen und behandeln. Springer, Berlin and New York (2011).
35. Neter, E., Brainin, E.: eHealth literacy: a marker for "digital divide" in health information. Rev. Health Care. 3, 145–151 (2012).
36. Tennant, B., Stellefson, M., Dodd, V., Chaney, B., Chaney, D., Paige, S., Alber, J.: eHealth Literacy and Web 2.0 Health Information Seeking Behaviors Among Baby Boomers and Older Adults. J. Med. Internet Res. 17, (2015).
37. Doelfs, G.: Elektronische Gesundheitsakte: Kampfansage aus Cupertino. Kma - Gesundheitswirtschaftsmagazin. 23, 62–63 (2018).
38. Prat, N., Comyn-Wattiau, I., Akoka, J.: A Taxonomy of Evaluation Methods for Information Systems Artifacts. J. Manag. Inf. Syst. 32, 229–267 (2015).

Digitalisierung in der Stressprävention – eine qualitative Interviewstudie zu Nutzenpotenzialen

Kim Janine Blankenhagel[1], Anne-Katrin Witte[1], Rüdiger Zarnekow[1]

[1] Technische Universität Berlin, Lehrstuhl für Informations- und Kommunikationsmanagement, Berlin, Deutschland
{k.blankenhagel,a.witte,ruediger.zarnekow}@tu-berlin.de

Abstract. Stress gilt als Verursacher zahlreicher Erkrankungen, wenn er nicht mit ausreichend Entspannung und Ruhe im Gleichgewicht steht. Dauerstress und Burnout nehmen stetig zu, weshalb die Notwendigkeit für Prävention und Stressmanagement steigt. Digitale Technologien ermöglichen neue Methoden präventiven Stressmanagements, allerdings sind die Nutzenpotenziale entsprechender Systeme weitgehend unerforscht. Diese Untersuchung schließt die Forschungslücke mittels einer qualitativen Interviewstudie. Dafür wurden 15 teilstandardisierte Interviews geführt, die vier verschiedene Perspektiven (Krankenkassen, Leistungserbringer, Privatwirtschaft, Nutzer) beinhalten. Aus ihnen wurden Nutzenpotenziale digitaler Stressmanagementsysteme aggregiert, die sich in die zwei entwickelten Kategorien *Verbesserung der medizinischen Versorgung* und *Verbesserung der Inanspruchnahme* unterteilen lassen. Die Interviews ergaben, dass digitale Stressmanagementsysteme diverse Nutzenpotenziale besitzen und einen Beitrag zur Gesundheitsversorgung leisten können. Sie können niedrigschwellig in Anspruch genommen werden, stärken die Eigenständigkeit und Selbsthilfe der Nutzer und bieten innovative Ansätze, übermäßigen Stress zu erkennen und adäquat darauf zu reagieren.

Keywords: digitales Stressmanagement, Prävention, Burnout, Nutzenpotenziale

1 Einleitung

Stress ist in vielen industrialisierten Ländern weit verbreitet [1] und kann zu diversen psychischen und physischen Erkrankungen führen [2]. Hohe Arbeitsbelastungen und gestiegene Komplexität im Arbeitsalltag führen zu steigenden Anforderungen an Berufstätige, was sich häufig durch erhöhte Stresslevel äußert [3]. Diese sind Vorstufen eines Burnouts und Stressmanagement ist neben ausreichender Bewegung einer der wichtigsten Faktoren, um den Ausbruch der Erkrankung zu vermeiden [4]. Die Partizipation an herkömmlichen präventiven Angeboten zum erfolgreichen Umgang mit Stress ist allerdings gering. Gleichzeitig verbessern sich die technologischen Möglichkeiten, mittels neuartiger Sensortechnik sowie Interpretation und Analyse großer Datenmengen die medizinische Versorgung zu verbessern [5].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Die gesundheitliche Versorgungskette lässt sich in die Bereiche Prävention, Diagnose, Therapie und Rückfallprävention unterteilen, wobei sich die vorliegende Untersuchung auf die präventive Phase bezieht. Der Forschungsschwerpunkt liegt demnach bei Systemen, die für gesunde Menschen entwickelt wurden und keinen therapeutischen oder krankheitsspezifischen Bezug aufweisen. Dieser Artikel umfasst eine Erhebung und Analyse von Nutzenpotenzialen mit folgender Forschungsfrage:

„Welche Nutzenpotenziale besitzen digitale Stressmanagementsysteme?“

Unter präventiven digitalen Stressmanagementsystemen (DSMS) werden im Rahmen dieser Erhebung Bewältigungsstrategien verstanden, die dem Erhalt der Gesundheit auf seelischer, körperlicher und sozialer Ebene dienen, das Ziel haben, Stress aufgrund arbeitsbezogener Belastungen in Richtung eines gesunden Gleichgewichts zu verringern und hauptsächlich auf digitalem Wege bereitgestellt werden (angelehnt an [6]). Sie können beispielsweise mittels Websites, Gesundheits-Applikationen auf dem Smartphone oder Wearables in Anspruch genommen werden. Der Begriff Nutzenpotenzial wird als Möglichkeit für eine im Sinne des Nutzers positive Veränderung definiert, die sich auf Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Patientenzufriedenheit bezieht (angelehnt an [7]).

Bisher sind digitale Präventionsangebote überwiegend unerforscht und die Nutzenpotenziale von DSMS in der Stressprävention weitgehend unbekannt [8]. Da zum jetzigen Zeitpunkt Einschätzungen, Deutungsmuster und Handlungsorientierungen sowie die Ermittlung individueller Perspektiven fehlen, ist im Rahmen dieser Untersuchung mit Blick auf ihren explorativen Charakter die Methodik qualitativer Interviews ausgewählt worden.

2 Einbettung in die Forschungslandschaft

Die Analyse relevanter Forschungsliteratur zum Thema digitales Stressmanagement zeigt auf, dass der Bereich trotz der enormen Brisanz bisher wenig betrachtet wurde. Der Großteil aller Studien untersucht therapeutische Systeme für bereits erkrankte Personen (zum Beispiel Menschen mit stressassoziierten Erkrankungen wie Depression oder Burnout) mit zum Großteil quantitativem Studiendesign. In der Literatur lassen sich einige wenige Studien finden, die Fragen der Wirksamkeit und moderierende Faktoren digitaler Stressmanagementsysteme thematisieren (zum Beispiel [9–11]). Sie untersuchen allerdings hauptsächlich medizinische Aspekte, während der vorliegende Artikel zusätzlich den Anwender und seine Bedürfnisse sowie gesamtgesellschaftliche Aspekte adressiert, sodass eine ganzheitliche Sicht entwickelt werden kann. Des Weiteren wurden verschiedene Perspektiven und Sichtweisen der relevanten Stakeholder einbezogen und verglichen, wodurch sich diese Analyse von bereits vorliegenden Studien abhebt.

In der Literatur werden neben Stressmanagement zwei weitere Begriffe zur Vermeidung eines Burnouts benannt. Die Definition von DSMS grenzt sich von dem in der Literatur thematisierten „Job Redesign“ insofern ab, als dass DSMS von den

Nutzern selbst angewendet werden, wohingegen Maßnahmen des Job Redesigns hauptsächlich von Seiten der Arbeitgeber umgesetzt werden und eine Änderung der Rahmenbedingungen fokussieren [12]. Maßnahmen des „Job-Craftings“, bei dem wie bei DSMS die Aktivität des Nutzers im Zentrum steht, sind ähnlich zu solchen von DSMS. Beim Job-Crafting liegen die Aktionen allerdings hauptsächlich in stark arbeitsbezogenen Interventionen [13], wobei DSMS in einem allgemeineren Ansatz stressbezogene Interventionen anbieten, die sich nicht alle zwingend auf den Themenbereich Arbeit beziehen müssen.

3 Methodik

Um die Nutzenpotenziale von DSMS zu eruieren, wurden im Frühjahr 2018 fünfzehn teilstandardisierte, leitfadengestützte Interviews durchgeführt. Mit dem Ziel eines ganzheitlichen Verständnisses wurden die vier zentralen Perspektiven der Krankenkassen (n = 3), der Leistungserbringer (Psychotherapeuten, medizinisches Personal; n = 4), der Privatwirtschaft (Anbieter von DSMS auf dem freien Markt; n = 4) und der Nutzer (n = 4) einbezogen. Insgesamt wurden 25 Interviewanfragen per E-Mail und eine Anfrage in einer Quantified-Self-Gruppe versendet, aus denen zehn telefonische und fünf persönliche Gespräche in deutscher Sprache resultierten.

Die Auswahl der Interviewpartner erfolgte nach Flickl [14] gemäß der Fallauswahl mittels eines qualitativen Stichprobenplans, um eine gezielte Auswahl besonders aussagekräftiger Fälle aufzunehmen. Konkret wurden die potenziellen Interviewpartner nach der (beruflichen) Erfahrung und Auseinandersetzung im Bereich der Stressprävention und der Nutzung digitaler Technologien zum Stressmanagement (mindestens zwei Jahre) rekrutiert. Die Voraussetzung zur Teilnahme am Interview war je nach Personengruppe entweder die Nutzung (Nutzer/Leistungserbringer) oder das Angebot (Privatwirtschaft/Krankenkassen) mindestens eines DSMS. Im Durchschnitt dauerte ein Interview 53 Minuten. Zu Beginn aller Interviews stellte der Interviewer sich und das Forschungsvorhaben vor und erläuterte Absicht sowie Struktur des Interviews. Mit dem Einverständnis der Interviewpartner wurden alle Gespräche aufgezeichnet, transkribiert und gemäß der qualitativen Inhaltsanalyse nach Mayring [15] mit einer Software für qualitative Analyse (MaxQDA) ausgewertet. Die identifizierten Nutzenpotenziale wurden codiert, aggregiert und analysiert. Abbildung 1 visualisiert das methodische Vorgehen, das in den folgenden Abschnitten präziser erläutert wird.

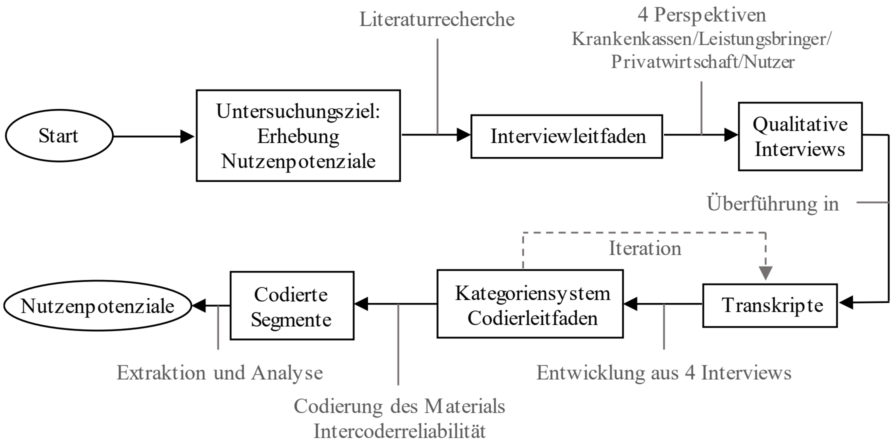


Abbildung 1. Methodisches Vorgehen

3.1 Interviewleitfaden

Der zugrundeliegende teilstandardisierte Interviewleitfaden basiert auf den etablierten Prinzipien von Döring und Bortz und beruht auf einer vorangegangenen Sichtung der Literatur [16]. Aus dieser wurden Hinweise zu Nutzenpotenzialen aus vorhandenen Konzepten und Theorien zu DSMS deduziert und daraus Interviewfragen abgeleitet. Zunächst wurden das Profil des Interviewpartners und entsprechende Erfahrungen erfragt. Im Anschluss wurden Erzählaufforderungen zum Aufbau, zur Funktionsweise und zur Interaktion mit dem vorhandenen bzw. genutzten DSMS gesetzt, um darauffolgend in die Erhebung der Nutzenpotenziale überzugehen. Neben den Nutzenpotenzialen wurden auch Nachteile und Grenzen von DSMS erfragt. Da sich der Wissensstand und die Perspektiven der befragten Akteursgruppen unterscheiden, wurden für jede der vier Personengruppen die Interviewleitfäden spezifisch angepasst. Allgemein sind die Fragen nach dem Prinzip der Offenheit formuliert [17], sodass die Interviewten ihrer Sachkenntnis und ihren Interessen entsprechend antworten konnten. Um sicherzustellen, dass der Leitfaden sowohl zeitlich als auch inhaltlich angemessen und zielführend ist, wurde er vorab einzeln in jeder Personengruppe getestet.

3.2 Kategoriensystem

Die Entwicklung eines Kategoriensystems wurde regelbasiert nach Mayring [15] durchgeführt. Zunächst wurde das gesamte Material gesichtet und einer der Autoren entwickelte induktiv in einem iterativen Prozess das Kategoriensystem. Des Weiteren wurde ein vorläufiger Codierleitfaden erarbeitet, in dem die Ausprägungen der einzelnen Kategorien definiert und um Ankerbeispiele erweitert wurden. Daneben beinhaltet der Codierleitfaden weitere Codierregeln. Demnach sind die entwickelten Kategorien disjunkt, erschöpfend und überlappungsfrei und gebildete Unterkategorien stellen Ausprägungen ihrer Oberkategorien dar. Es wurden nur zusammenhängende

Textstellen mit mindestens drei Worten und maximal fünf Sätzen codiert. Nach einer Probecodierung von vier Interviews wurde eine endgültige Fassung des Codierleitfadens festgelegt. Um die Intercoderreliabilität sicherzustellen, codierte einer der anderen Autoren einen Teil des Interviewmaterials mit einem Kohens-Cappa-Koeffizienten von 0,81. Nach gemeinsamer Diskussion der unklaren Zuordnungen wurden die restlichen Interviews codiert. Es sind zwei Hauptkategorien (*Verbesserung der medizinischen Prävention* und *Verbesserung der Inanspruchnahme*) entstanden, die Abbildung 2 entnommen werden können. Sie veranschaulicht überdies die acht entwickelten Unterkategorien.

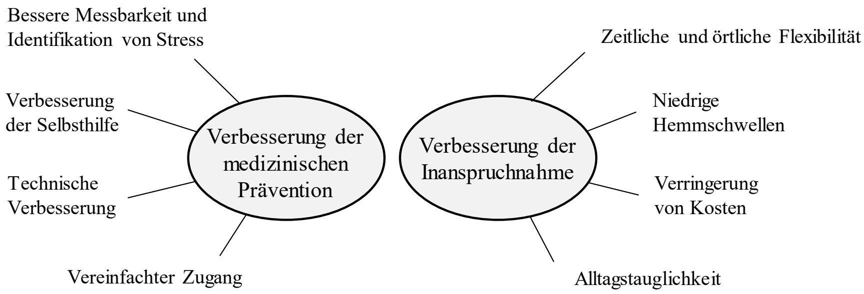


Abbildung 2. Kategoriensystem

4 Nutzenpotenziale von DSMS

Die erhobenen Interviews ergeben Nutzenpotenziale in den zwei Bereichen medizinische Prävention und Inanspruchnahme. Mit einem Anteil von 61 % und 99 Textstellen (Codes) wurde in Summe die Kategorie *Verbesserung der medizinischen Prävention* am häufigsten thematisiert. An zweiter Stelle folgen Aussagen zu besseren Bedingungen der Inanspruchnahme (63 Codes, 39 %). Betrachtet man die Verteilung in den einzelnen befragten Gruppen getrennt, so spiegelt sich die oben genannte Reihenfolge bei den Anbietern der Privatwirtschaft und bei den Leistungserbringern wieder. Hingegen nennen die Nutzer und die Krankenkassen Nutzenpotenziale im Zusammenhang mit der Verbesserung der medizinischen Prävention und bezüglich der Inanspruchnahme nahezu gleich häufig mit einem leichten Schwerpunkt auf der Verbesserung der Inanspruchnahme. An dieser Verteilung lässt sich erkennen, dass die Nutzer und die Krankenkassen den Nutzen von DSMS gleichermaßen in medizinischen wie in nicht-medizinischen Gesichtspunkten (verbesserte Bedingungen für die Inanspruchnahme) sehen. Für die Privatwirtschaft und die Leistungserbringer ist der medizinische Fokus ausgeprägter, was mit ihrer beruflichen Ausrichtung und Erfahrung einhergeht und übereinstimmt. Abbildung 3 illustriert die Verteilung der Codes.

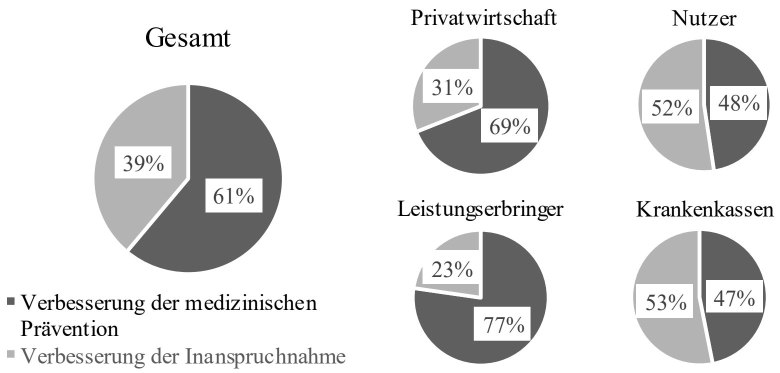


Abbildung 3. Verteilung der Codes

4.1 Verbesserung der medizinischen Prävention

Dieser Abschnitt thematisiert die von den Interviewpartnern genannten Nutzenpotenziale beim Einsatz von DSMS, die zu einer Verbesserung der medizinischen Prävention führen. Eingeschlossen wurden alle Aussagen der Interviewpartner, die sich auf den Umgang, die Messung und Bewertung von Stress beziehen, Empfehlungen, Übungen oder andere medizinische Anregungen bezüglich digitalem Stressmanagement beinhalten oder die Interaktion mit tragbaren Computer- und Internettechnologien thematisieren, die einen direkten Einfluss auf die medizinische Prävention von Stress besitzt. Zusammenfassend werden die bessere Messbarkeit und Identifikation von Stress (42 Codes, 42 %), die verbesserte Selbsthilfe (19 Codes, 19 %), technische Verbesserungen (22 Codes, 22 %) und der vereinfachte Zugang (16 Codes, 16 %) genannt.

Bessere Messbarkeit und Identifikation von Stress. Die Unterkategorie *Bessere Messbarkeit und Identifikation von Stress* ist innerhalb der Kategorie *Verbesserung der medizinischen Prävention* am stärksten thematisiert. Dies zeigt, dass die Funktionalität der Stressmessung bei DSMS weitreichende Nutzenpotenziale mit sich bringt. Bis auf die Krankenkassen diskutieren diese Thematik alle befragten Akteursgruppen ausführlich. Das zurückhaltende Verhalten der Krankenkassen könnte damit erklärt werden, dass die DSMS der Krankenkassen hauptsächlich websitebasierte Programme darstellen, die über keine Stress messenden Funktionen verfügen.

Die Leistungserbringer erklären, dass Stress ein unspezifisches und abstraktes Erscheinungsbild ist und erst dann gesundheitsschädlich oder sogar krankhaft wird, wenn er häufig und langanhaltend auftritt. Die Messung von Stress steht daher im Zentrum der Aufmerksamkeit. Da Stress zu körperlichen Reaktionen wie beispielsweise vermehrter Schweißbildung, flacher Atmung oder Verspannungen führt, können durch Messung entsprechender Parameter Hinweise auf aktuelle Stresslevel gewonnen werden. Die Leistungserbringer und Anbieter aus der Privatwirtschaft benennen die Messung der Herzratenvariabilität, des Hautleitwiderstands, der

Atemfrequenz oder muskulärer Spannungen (zum Beispiel im Nacken) als geeignete Parameter zur Bestimmung von Stressbelastungen. Dafür bieten sich insbesondere Wearables an.

Wenn die gemessenen Größen vom DSMS in ein für den Nutzer verständliches Stresslevel übersetzt werden, wird er in die Lage versetzt, eigene Stressbelastungen zu erkennen und auszuwerten. Ursache- und Wirkungsbeziehungen können klar dargestellt werden, was einen hohen motivationalen Mehrwert mit sich bringt. Die Nutzer berichten davon, dass ein Stress messendes Gerät schnell eine Ankerfunktion übernimmt und sie es nach einigen Nutzungszyklen - durch Konditionierung - bereits mit Entspannung und Stressabbau assoziieren. Zu Beginn der Stressprävention ist es besonders gewinnbringend, wenn eine Auswertung der Stresslevel detailliert im Zeitverlauf dargestellt wird, da der Nutzer seine potenziell hohen Stresswerte mit entsprechenden Situationen und Ereignissen in Verbindung setzen kann. So wird es dem Nutzer erleichtert, seine individuellen externen Stressoren zu identifizieren. An dieser Stelle erkennen insbesondere die Anbieter der Privatwirtschaft, die Nutzer und die Leistungserbringer beträchtliche Nutzenpotenziale und sehen einen klaren Mehrwert gegenüber herkömmlichen Präventionsmaßnahmen (wie beispielsweise Informationsveranstaltungen oder routinemäßige Gesundheits-Checks beim Arzt). Des Weiteren ermöglicht die selbstständige Messung des Stresslevels dem Nutzer, geeignete Übungen zur Stressreduktion abzuleiten und deren Erfolg zu bewerten. Die Interviewpartner betonen den Vorteil tragbarer Internettechnologien insofern, als mit ihnen Stressmessung auch im alltäglichen Leben unabhängig von einem Arztbesuch möglich ist.

Insgesamt bieten DSMS neuartige Möglichkeiten, unabhängig und alltagsnah Stresswerte zu messen und zu dokumentieren. Diese Funktionen haben Potenzial, die Diagnostik von Stress oder eines drohenden Burnouts zu verbessern, und stellen eine fundierte Basis dar, um Anti-Stress-Maßnahmen auszuwählen und durchzuführen.

Verbesserung der Selbsthilfe. In der Unterkategorie *Verbesserung der Selbsthilfe* finden sich Aussagen aller interviewten Akteursgruppen, wobei sich die Leistungserbringer mit ihrem medizinisch-therapeutischen Wissen besonders intensiv einbringen.

Aus Sicht der Interviewpartner können DSMS in drei Bereichen Unterstützung zur Selbsthilfe leisten. Zunächst stellen sie Material zur Wissensvermittlung zur Verfügung, in dem klassischerweise Inhalte zu Stressarten, Warnsignalen, Symptomen und der Wirksamkeit (präventiver) Maßnahmen vermittelt werden. Laut der Anbieter der Privatwirtschaft, der Leistungserbringer und der Krankenkassen lässt sich dieser Bereich sehr gut an digitale Dienste abgeben, da in der Wissensvermittlung weder individuelle Interaktionen noch persönliche Beziehungen vonnöten sind. Auf diese Weise entsteht Zeitersparnis und die Nutzer können ihren Fokus auf weitere Übungen legen. Der zweite Bereich beschreibt Interventionen reflektiver Art, die häufig achtsamkeitsbasiert sind und Körpergefühl sowie -wahrnehmung schulen. Mit dem Ziel einer verbesserten Selbstbeobachtung erlernen die Nutzer autonom, ihren Gesundheitszustand besser einzuschätzen. Durch die begleitende Messung der Stresslevel können die Nutzer außerhalb des ärztlichen Umfelds relevante Parameter aufnehmen und diese eigenständig mit den Erkenntnissen der Selbstbeobachtung abgleichen. Des Weiteren umfassen achtsamkeitsbasierte Übungen kurz- bis mittelfristige Bewältigungsstrategien (wie beispielsweise Atemübungen), die den Nutzern in stressigen Phasen unverzüglich helfen können, Stress abzubauen.

Der dritte Bereich besteht aus Interventionen, die Übungen zum Aufbau von Anti-Stress-Ressourcen zur Verfügung stellen und hauptsächlich der langfristigen Stressbewältigung dienen. Die diesbezügliche Eignung von DSMS wird von den Interviewpartnern rege diskutiert. Die Anbieter in der Privatwirtschaft und die Nutzer erkennen Nutzenpotenziale hinsichtlich der Entscheidung, welche Ressourcen für ausreichend Erholung im Alltag sorgen können. Demnach sehen sie Vorteile in allen drei genannten Bereichen, auch wenn die dritte Ebene klassischerweise im Rahmen einer Therapie thematisiert und zwischenmenschlich bearbeitet wird. Da eine Therapie erst nach Ausbruch einer Erkrankung ansetzt, bieten DSMS die Möglichkeit, schon vorher eine Präventionskultur zu etablieren und gleichzeitig langfristige, individuell passende Stressbewältigungsstrategien für den Nutzer zu eruieren.

Technische Verbesserungen. Die Unterkategorie *Technische Verbesserungen* ist stark geprägt von Aussagen der Privatwirtschaft, für die die Entwicklung und technische Umsetzung von DSMS naturgemäß im Zentrum stehen. Auch die Leistungserbringer bringen sich rege ein.

Ein Vorteil tragbarer Computer- bzw. Internettechnologien liegt unter anderem in permanent durchführbarer Stressmessung. Folglich kann sofortiges Feedback durch Hinweise und Warnungen an den Nutzer gesendet werden, sodass Maßnahmen zum kurzfristigen Stressabbau unmittelbar getroffen und die stressige Situation zeitnah gelöst werden kann. Diese dauerhafte Verfügbarkeit wird von den Interviewpartnern als Schlüsselkomponente für aktives und wirksames Stressmanagement eingestuft.

Die Stressmessung und die Erhebung weiterer Daten wie beispielsweise Selbsteinschätzungen des Nutzers bieten eine solide Datengrundlage, um fundiert Einschätzungen des Gesundheitszustands zu treffen. Daneben lassen die Daten Rückschlüsse auf die Wirksamkeit durchgeführter Interventionen und Übungen zu. Neben der Verwendung der Datenmenge für ein wirksames Stressmanagement könnte diese mit dem Einverständnis des Nutzers auch für bevölkerungsweite Analysen oder für solche von Teilpopulationen genutzt werden, um Vorhersagen und Annahmen über Krankheitsverläufe (gleichartiger Patientengruppen) zu verbessern. Dies birgt das Potenzial einer stärkeren Evidenz wissenschaftlicher Forschung.

Eine maßgeschneiderte Bereitstellung für den individuellen Nutzer stellt aus Sicht der Interviewpartner ein sehr bedeutendes Nutzenpotenzial dar. Die Leistungserbringer und die Privatwirtschaft konkretisieren, dass Individualität technisch mithilfe künstlicher Intelligenz geeignet realisiert werden kann. Es werden lernende Systeme favorisiert, die auf Basis von Algorithmen das Profil und die Merkmale eines Nutzers mit der Zeit erlernen und somit zu einem sehr individuellen DSMS führen, das Stresssituationen sowie wirksame Maßnahmen nutzerbezogen eruiert. Durch die Erstellung nutzerspezifischer Empfehlungen und Zusammenfassungen entstehen präzise und gewinnbringende Rückmeldungen und Interaktionen zwischen DSMS und Nutzer. Auch die automatische Erkennung von Stimmlagen, Gemütszustandsveränderungen sowie Abweichungen vom normalen Muster zählt zu den Funktionalitäten, die durch den Einsatz künstlicher Intelligenz vermehrt zur Verfügung gestellt werden können. Die Leistungserbringer betonen, dass hinter jeder Stressbelastung oder arbeitsbezogenen Depression ein sehr individueller Verlauf steckt, der bei nicht-individualisierten DSMS keine Berücksichtigung findet. Daher sollten die Themenbereiche Individualisierung und künstliche Intelligenz verstärkt fokussiert werden.

Vereinfachter Zugang. Digitale Technologien besitzen den Vorteil eines einfachen Zugangs für einen Großteil der Bevölkerung. Dieser Aspekt wird insbesondere von den Krankenkassen hervorgehoben. Durch einfache und schnelle Zugangsmöglichkeiten wie das Aufrufen einer Website oder das Herunterladen einer Applikation auf das Smartphone kann eine große Anzahl an Personen mit nur geringem Aufwand zeitunabhängig beginnen, mit einem DSMS zu arbeiten. Dies ist besonders für Menschen in strukturschwachen Regionen mit wenig Zugang zu medizinischer Unterstützung von Vorteil. Dank der guten Skalierbarkeit können eine große Zielgruppe und auch neue Personengruppen, die von herkömmlichen Präventionsangeboten nicht angesprochen werden, erreicht werden. Dazu zählen insbesondere junge Menschen und technikaffine Personen.

4.2 Verbesserung der Inanspruchnahme

Der Einsatz digitaler Technologien bewirkt einige Nutzenpotenziale, die sich nicht direkt auf die medizinische Versorgung beziehen. Dieser Abschnitt thematisiert die von den Interviewpartnern genannten Nutzenpotenziale im Einsatz von DSMS bezüglich der Verbesserung der Bedingungen der Inanspruchnahme. Eingeschlossen wurden alle Aussagen der Interviewpartner, die sich auf die Art und Weise der Nutzung des Systems sowie auf die Rahmenbedingungen der Nutzung und dessen (nicht-medizinische) Auswirkungen beziehen. Zusammenfassend werden zeitliche und örtliche Flexibilität (20 Codes, 32 %), eine sehr gute Alltagstauglichkeit (16 Codes, 25 %), niedrige Hemmschwellen (16 Codes, 25 %) sowie eine Verringerung von Kosten (11 Codes, 18 %) genannt.

Zeitliche und örtliche Flexibilität. Die Unterkategorie *Zeitliche und örtliche Flexibilität* ist geprägt von Äußerungen der Nutzer und der Krankenkassen. Aufgrund der Schnelllebigkeit im (beruflichen) Alltag ist Zeit bei stressgeplagten Menschen oder bei Personen mit ausgeprägter Arbeitsbelastung häufig ein sehr kritischer Faktor. Daher sind kurze Bearbeitungszeiten (die Interviewpartner empfehlen 5–15 Minuten pro Tag) und zeitliche Flexibilität Parameter, die für viele Personen (insbesondere bezüglich der genannten Zielgruppe) von besonderer Bedeutung sind. Da die Nutzung eines DSMS weder mit Fahrtzeiten noch mit festen Terminen verbunden ist, bietet sie neue Freiräume durch flexible Nutzung. Auch die Ortsunabhängigkeit ist gerade für Berufstätige von Relevanz. Die mögliche Verwendung eines DSMS am Arbeitsplatz oder in gewohnten Umgebungen, wie dem eigenen Zuhause, eröffnet diverse Vorteile.

Alltagstauglichkeit und niedrige Hemmschwellen. Alle Interviewpartner sehen in der Nutzung von DSMS Vorteile aufgrund der hohen Alltagstauglichkeit. Da DSMS häufig auf mobilen Endgeräten wie Smartphones oder Wearables verfügbar sind, lassen sie sich nach den Erfahrungen der Nutzer sehr gut mit sich führen und in den Alltag integrieren. Demnach ist es möglich, während des Tages kurzfristig auf Hinweise des DSMS zu reagieren. Außerdem führt eine intelligente Steuerung von Hinweisen und Handlungsempfehlungen (beispielsweise keine Störungen bei einem Aufenthalt in einem bestimmten Konferenzraum) dazu, dass die Interaktion mit einem DSMS an situative Gegebenheiten angepasst ist und daher die Integration in den Alltag weiter verbessert wird.

Digitale Dienste wie DSMS können anonym und mit nur geringem Aufwand in Anspruch genommen werden. Dies senkt unter den Nutzern die Hemmschwelle, entsprechende Systeme auch tatsächlich zu gebrauchen. Die Leistungserbringer berichten aus ihrer Erfahrung, dass im Bereich psychischer Erkrankungen oder eines Burnouts Anonymität und Unverbindlichkeit sehr häufig wichtige Voraussetzungen dafür darstellen, dass sich Risikopatienten oder auch Menschen mit Symptomen überhaupt dafür entscheiden, Unterstützung anzunehmen. Der geringe Aufwand, mithilfe von DSMS Maßnahmen zur Stressbewältigung zu treffen, und die damit verbundene Anonymität und Unverbindlichkeit führen aus Sicht der Interviewpartner zu klaren Vorteilen und einer allgemein höheren Beteiligung.

Verringerung von Kosten. Die Privatwirtschaft, Leistungserbringer und Krankenkassen sind der Überzeugung, dass der Einsatz von DSMS langfristig zu Kosteneinsparungen im Gesundheitssystem führt. Die Kategorie *Verringerung von Kosten* ist in starkem Maße von den Aussagen der Krankenkassen geprägt. Einsparpotenziale werden zum einen darin gesehen, dass infolge erfolgreicher Prävention die Zahl der Burnout-Erkrankungen und damit die Behandlungskosten und Krankenhausaufenthalte zurückgehen. Des Weiteren bieten DSMS die Möglichkeit, im Falle des Ausbruchs einer Burnout-Erkrankung einige geeignete Therapiebestandteile wie beispielsweise die Psychoedukation, an digitale Systeme abzugeben und damit Therapiekosten bei gleichbleibender Qualität einzusparen. Aufgrund der guten Skalierbarkeit sind DSMS ökonomisch effizient. Neben der Verringerung der Kosten für das Gesundheitssystem führt aus wirtschaftlicher Sicht die allgemeine Verringerung krankheitsbedingter Ausfälle zu einer gesteigerten Produktivität. Die Forschung könnte vom Einsatz von DSMS durch eine schnelle und kosteneffektive Sammlung von Daten profitieren, die sonst traditionell mit Fragebögen erhoben werden.

Tabelle 1 listet die genannten Nutzenpotenziale auf und gibt zusätzlich an, ob sich diese konkret auf Stress und Burnout beziehen oder auch auf andere strukturgleiche Anwendungsfälle übertragbar sein könnten. Diese Beurteilung stammt aus den Erkenntnissen des Interviewmaterials und wurde um eine gemeinsame Diskussion der Autoren ergänzt. Des Weiteren ist durch Kreise gekennzeichnet, wie häufig das jeweilige Nutzenpotenzial von den verschiedenen Akteuren genannt wurde.

Tabelle 1. Nutzenpotenziale von DSMS

Nutzenpotenziale	*PW*	*N*	*LE*	*KK*
Bessere Messbarkeit und Identifikation von Stress				
Arzt-unabhängige und alltagsnahe Messung von Stress	●	◑	●	◑
Bessere Diagnostik und Überwachung von Stress	●	◑	●	○
Auswertung individueller Stressbelastungen im Zeitverlauf zur Identifikation externer Stressoren	●	●	●	○
Verbesserung der Selbsthilfe				
Kurzweilige, digitale Wissensvermittlung	●	◑	●	●
Interventionen reflektiver Art für kurz- bis mittelfristige Stressbewältigungsstrategien	●	◑	●	◑
Interventionen zum Aufbau von Anti-Stress-Ressourcen zur langfristigen Stressbewältigung	●	●	◑	◑
Technische Verbesserung				
Sofortiges Feedback (bspw. in Stresssituationen) vom DSMS durch Hinweise und Warnungen	●	●	●	◑
Bessere Datengrundlage (Gesundheitszustand/ Wirksamkeit/ Evidenz)	●	○	●	◑
Einsatz künstlicher Intelligenz durch lernende Systeme	●	◑	●	◑
Vereinfachter Zugang				
Einfache und schnelle Zugangsmöglichkeiten; sehr gute Skalierbarkeit	◑	○	◑	●
Zeitliche und örtliche Flexibilität				
Zeitliche Flexibilität, keine festen Termine oder Fahrtzeiten; Nutzung an diversen Orten möglich	◑	●	◑	●
Alltagstauglichkeit				
Gute Integration in den Alltag; ständige Stressmessung	●	●	●	●
Niedrige Hemmschwellen				
Anonyme und unverbindliche Nutzung	◑	●	●	●
Verringerung von Kosten				
Kosteneinsparungen aufgrund geringerer Anzahl an Burnout-Erkrankungen, geringeren Behandlungskosten, höherer Produktivität	◑	○	◑	●

▭ Burnoutspezifisch ● ≥ 50% ◑ < 50% ○ Keine Nennung
PW = Privatwirtschaft N = Nutzer LE = Leistungserbringer KK = Krankenkassen

5 Grenzen von DSMS

Mit insgesamt 37 Textstellen adressieren die Interviewpartner neben den zahlreichen Nutzenpotenzialen ebenfalls Grenzen und Nachteile von DSMS. Alle Akteursgruppen benennen fehlende wissenschaftliche Fundierung sowie Evidenz und die damit einhergehende mangelnde Sicherstellung von Qualitätsansprüchen. Insbesondere für die Nutzer ist nicht ersichtlich, welche auf dem Markt angebotenen DSMS eine hohe Qualität aufweisen.

Auch datenschutzrechtliche Risiken werden von den Interviewpartnern häufig genannt. Mangelnde Sicherstellung von Anonymität, unsichere Datenübertragung mit möglichen Missbrauchsfolgen sowie fehlende Aufklärung und fehlende Datenhoheit werden DSMS nachteilig zugeschrieben.

Die Anbieter aus der Privatwirtschaft, die Krankenkassen und die Leistungserbringer diskutieren hohe vorherrschende Abbruchquoten bei DSMS. Insbesondere bei reinen Selbstmanagement-Programmen, die unbegleitet sind, scheinen Motivation und Durchhaltevermögen bei einigen Nutzern problematisch zu sein. Die Experten führen dies auf die naturgemäße Unverbindlichkeit von DSMS und auf das Fehlen einer menschlichen bzw. sozialen Komponente zurück. An dieser Stelle weisen die Leistungserbringer und die Krankenkassen darauf hin, dass aufgrund der individuellen Umgebung, der Berufswelt und konkreten Arbeitssituation mit ihren jeweiligen spezifischen Merkmalen und Anforderungen digitale Systeme den Bereich der langfristigen und strategischen Stressbewältigung ohne menschlichen Einbezug nicht ausreichend abbilden können. Daher sehen sie für diesen Teilbereich der Stressprävention die Integration medizinischen Fachpersonals als notwendig an und erkennen Grenzen der Internettechnologie.

Allgemein entstehen durch die Nutzung von DSMS Abhängigkeiten zu einer Internetanbindung, zu einem Stress messenden Gerät bzw. einem Smartphone. Die Leistungserbringer ergänzen, dass durch dessen ständige Verfügbarkeit eine klare Trennung zwischen Alltag und Stressreduktion bzw. Arbeit und Stressreduktion fehlt. Herkömmliche Präventionsmaßnahmen sind oft mit einer Anreise verbunden, wodurch in der Regel eine bessere Konzentration für die einzelnen Interventionen aufgebracht werden kann. Tabelle 2 listet die genannten Grenzen auf und gibt zusätzlich an, ob sich diese konkret auf Stress und Burnout beziehen oder auch auf andere strukturgleiche Anwendungsfälle übertragbar sein könnten.

Tabelle 2. Grenzen von DSMS

Grenzen	*PW*	*N*	*LE*	*KK*
Mangelnde Sicherstellung von Qualitätsansprüchen	●	●	●	●
Nichteinhaltung datenschutzrechtlicher Ansprüche	●	◑	●	◑
Hohe Abbruchquoten	◑	○	●	◑
Abhängigkeiten von Internetanbindungen oder Stress messenden Geräten	○	◑	●	○
Fehlende Abgrenzung der Stressreduktion vom Alltag	○	◑	●	○

▭ Burnoutspezifisch ● ≥ 50% ◑ < 50% ○ Keine Nennung

PW = Privatwirtschaft N = Nutzer LE = Leistungserbringer KK = Krankenkassen

6 Diskussion

Die Analyse des Interviewmaterials lässt den Schluss zu, dass DSMS in der Lage sind, sowohl kognitive, emotionale als auch muskuläre Ebenen der Stressreduktion zu adressieren. Die Bereiche der Wissensvermittlung und der Unterstützung der Achtsamkeit sowie Körperwahrnehmung verfügen aus Sicht der Interviewpartner über besonders große Nutzenpotenziale, die vor allem auf die neu entstehenden Möglichkeiten der alltäglichen und dauerhaften Stressmessung zurückgeführt werden. Der Einsatz von Wearables macht Symptome auf medizinischer Grundlage erkennbar und lässt Rückschlüsse auf geeignete Stressreduktionsmaßnahmen zu. Dem stehen DSMS auf Basis selbst wahrgenommener Reaktionen gegenüber, die über keine Stress messenden Funktionalitäten verfügen. Da diese in der Regel weniger individuell sind und meist nur generelle Übungen und Hinweise vermitteln, werden sie im Vergleich als nachteilig angesehen. Der Einbezug einer Stress messenden Technologie wird von den Interviewpartnern klar gewünscht. Der Großteil der in dieser Untersuchung eruierten Nutzenpotenziale im Bereich der medizinischen Prävention entsteht erst durch das Vorhandensein alltagsnaher Stressmessung (beispielsweise bessere Diagnostik/Überwachung von Stress, sofortiges Feedback, Einsatz künstlicher Intelligenz etc.). Ob allerdings sofortiges Feedback mittels Hinweise oder Warnungen in Stresssituationen auch dazu führen könnte, dass eine Stressreaktion noch weiter verschärft wird, bleibt unklar und sollte in weiteren Forschungsvorhaben adressiert werden.

Des Weiteren ist es fragwürdig, inwiefern DSMS auch in der langfristigen Stressbewältigung und beim Aufbau von Anti-Stress-Ressourcen Nutzen bringen. Auf der einen Seite sind stressauslösende Situationen derart komplex und individuell verschieden, dass für strategische Stressbewältigungsstrategien eine rein digitale Bearbeitung ohne menschlichen Einbezug nicht ausreichend ist. Auf der anderen Seite ist denkbar, dass die Weiterentwicklung künstlicher Intelligenz derart individuelle Systeme hervorbringen kann, dass arbeitsbezogene Ausgangsbedingungen nach therapeutischen Verfahren erfolgreich analysiert und interpretiert werden können und damit das DSMS dem Nutzer maßgeschneiderte Anregungen vermitteln kann.

Im Vergleich lässt sich feststellen, dass die in dieser Untersuchung ermittelten nicht-burnoutspezifischen Vorteile digitaler Technologien wie beispielsweise vereinfachter Zugang, Anonymität oder zeitliche Flexibilität mit den in der Literatur bekannten Nutzenpotenzialen übereinstimmen. Hingegen sind Kenntnisse über die erarbeiteten burnoutspezifischen Nutzenpotenziale von DSMS in der Literatur weniger verbreitet. Dies zeigt, dass zur erfolgreichen Stressprävention und damit zur Vermeidung eines Burnouts einige Besonderheiten gelten, die in der Entwicklung entsprechender DSMS Berücksichtigung finden sollten. Da typischerweise Menschen mit einer intensiven arbeitsbezogenen Belastung und mit wenig Erholungsmöglichkeiten burnoutgefährdet sind, sollten DSMS mit möglichst kurzen Bearbeitungszeiten bedienbar sein. Sie adressieren somit Schnelllebigkeit und Zeitdruck, sodass auch stark stressgeplagte Menschen zeitliche Freiräume für die Nutzung realisieren können. Des Weiteren kann den derzeit vorliegenden hohen Abbruchquoten mittels Gamification und Motivationselementen begegnet werden.

Zukünftige Forschungsbereiche werden von den Autoren darin gesehen, die neu erhobenen Nutzenpotenziale anhand realer Anwendungsbeispiele zu untersuchen und quantitativ zu validieren. Durch den Einbezug der Perspektive von Arbeitgebern, die als ein Teil der stressauslösenden Quellen eines Burnouts angesehen werden, können eine Erweiterung der Nutzenpotenziale und ein Vergleich mit wirtschaftlichen Anforderungen erfolgen.

7 Limitationen

Die subjektive Verwendung sprachlichen Materials kann zu Interpretationsdifferenzen führen. Daher lassen sich subjektive Einflüsse aufgrund der persönlichen Erfahrung und des Wissensstands der Autoren im Rahmen der Interviewführung und bei der Auswertung der erhobenen Interviews nicht ausschließen. Durch Cross-Validierung wurde dieses Risiko minimiert.

Des Weiteren kann es aufgrund der unterschiedlichen Interviewformate (telefonisch oder persönlich) zu situativen Unterschieden gekommen sein.

Aufgrund der Voraussetzung, nur am Interview teilnehmen zu können, wenn ein DSMS bereits genutzt bzw. angeboten wird, könnten einseitig verzerrte Nutzenpotenziale erhoben worden sein. Somit ist eine Generalisierbarkeit der Aussagen auf die gesamte Bevölkerung fraglich. Dies betrifft insbesondere die Perspektive der Nutzer: die frühe Verwendung neuartiger Technologien lässt eine grundsätzliche Offenheit und Präferenz digitaler Systeme vermuten.

Alle Interviewpartner sind Teilnehmer im deutschen Gesundheitssystem, das durch deutsche und europäische Gesetze und Vorschriften wesentlich geprägt ist. Daher ist eine Übertragung auf nicht-europäische Länder zunächst nicht uneingeschränkt möglich.

8 Fazit

Zusammenfassend bestätigen die geführten Interviews, dass DSMS diverse Nutzenpotenziale besitzen und einen Beitrag zur Gesundheitsversorgung leisten können. Mittels digitaler Technologien zur Stressmessung entstehen neue Möglichkeiten, übermäßigen Stress zu erkennen und darauf zu reagieren. DSMS können niedrigschwellig in Anspruch genommen und die Eigenständigkeit und Selbsthilfe der Nutzer gestärkt werden. Die Digitalisierung der Prävention bringt allerdings auch Nachteile mit sich, unter die insbesondere datenschutzrechtliche Risiken, hohe Abbruchquoten und mangelnde Sicherstellung von Qualitätsansprüchen fallen.

Referenzen

1. Hasson, D., Anderberg, U.M., Theorell, T., Arnetz, B.B.: Psychophysiological effects of a web-based stress management system. A prospective, randomized controlled intervention study of IT and media workers [ISRCTN54254861]. BMC Public Health 5, 78 (2005)

2. Heber, E., Ebert, D.D., Lehr, D., Cuijpers, P., Berking, M., Nobis, S., Riper, H.: The Benefit of Web- and Computer-Based Interventions for Stress. A Systematic Review and Meta-Analysis. J. Med. Internet Res. 19, e32 (2017)
3. Möltner, H., Leve, J., Esch, T.: Burnout-Prävention und mobile Achtsamkeit. Evaluation eines appbasierten Gesundheitstrainings bei Berufstätigen. Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany)) 80, 295–300 (2018)
4. Walter, U., Krugmann, C.S., Plaumann, M.: Burn-out wirksam prävenieren? Ein systematischer Review zur Effektivität individuumbezogener und kombinierter Ansätze. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 55, 172–182 (2012)
5. Pisani, A.R., Wyman, P.A., Mohr, D.C., Perrino, T., Gallo, C., Villamar, J., Kendziora, K., Howe, G.W., Sloboda, Z., Brown, C.H.: Human Subjects Protection and Technology in Prevention Science. Selected Opportunities and Challenges. Prev. Sci. 17, 765–778 (2016)
6. Günthner, A., Batra, A.: Stressmanagement als Burn-out-Prophylaxe. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 55, 183–189 (2012)
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (ed.): Allgemeine Methoden Version 5.0 vom 10.07.2017, https://www.iqwig.de/de/methoden/methodenpapier.3020.html (Accessed: 18.08.2018)
8. Blankenhagel, K.J., Zarnekow, R.: Digitalisierung in der psychotherapeutischen Versorgung - ein Literatur-Review zum Status quo internet- und mobilbasierter Versorgungsprogramme und Ableitung einer Klassifizierung. In: MKWI 2018 - Multikonferenz der Wirtschaftsinformatik 2018, pp.683–694, Leuphana Universität, Lüneburg (2018)
9. Billings, D.W., Cook, R.F., Hendrickson, A., Dove, D.C.: A web-based approach to managing stress and mood disorders in the workforce. J. Occup. Environ. Med. 50, 960–968 (2008)
10. Ebert, D.D., Lehr, D., Heber, E., Riper, H., Cuijpers, P., Berking, M.: Internet- and mobile-based stress management for employees with adherence-focused guidance. Efficacy and mechanism of change. Scand. J. Work Environ. Health 42, 382–394 (2016)
11. Heber, E., Lehr, D., Ebert, D.D., Berking, M., Riper, H.: Web-Based and Mobile Stress Management Intervention for Employees. A Randomized Controlled Trial. J. Med. Internet Res. 18, e21 (2016)
12. Tims, M., Bakker, A.B., Derks, D.: The impact of job crafting on job demands, job resources, and well-being. J. Occup. Health Psychol. 18, 230–240 (2013)
13. Bakker, A.B., Demerouti, E.: The Job Demands-Resources model: state of the art. J. Manage. Psychol. 22, 309–328 (2007)
14. Flick, U.: Qualitative Sozialforschung. Eine Einführung. Rowohlts Enzyklopädie im Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg (2017)
15. Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz Verlagsgruppe, Weinheim, Basel (2010)
16. Döring, N., Bortz, J. (eds.): Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Springer, Berlin, Heidelberg (2016)
17. Lamnek, S., Krell, C. (eds.): Qualitative Sozialforschung. Mit Online-Materialien. Beltz, Weinheim, Basel (2016)

User Dynamics in Mental Health Forums – A Sentiment Analysis Perspective

Elena Davcheva, Martin Adam, Alexander Benlian

Technical University Darmstadt, Information Systems & E-Services, Darmstadt, Germany
{davcheva,adam,benlian}@ise.tu-darmstadt.de

Abstract. Individuals around the world in need of mental healthcare do not find adequate treatment because of lacking resources. Since the necessary support can often not be provided directly, many turn to the Internet for assistance, whereby mental health forums have evolved into an important medium for millions of users to share experiences. Information Systems research lacks empirical evidence to analyze how health forums influence users' moods. This paper addresses the research gap by conducting sentiment analysis on a large dataset of user posts from three leading English-language forums. The goal of this study is to shed light on the mood effects of mental health forum participation, as well as to better understand user roles. The results of our exploratory study show that sentiment scores develop either positively or negatively depending on the condition. We additionally investigate and report on user forum roles.

Keywords: Mental Health, Sentiment Analysis, Big Data, Forums, Natural Language Processing

1 Introduction

Mental disorders are defined as "a combination of abnormal thoughts, perceptions, emotions, behavior and relationships with others" [1]. The term comprises depression, bipolar disorder, schizophrenia, dementia, and developmental disorders such as autism. An estimated 300 million people are affected by depression alone (worldwide) and 15% of people aged 60 and over suffer from one or more disorders [2]. What is more, the world population is aging rapidly. Between 2015 and 2050, the proportion over 60 years will nearly double, from 12% to 22% [2]. Health systems have not yet adequately answered the growing burden. Between 35% and 85% of affected individuals receive no official treatment [3]. In many countries, less than one psychiatrist per 100,000 people is available [4]. Moreover, the fear of stigma discourages people from seeking assistance [5]. Even if willing, those affected often cannot afford the medical treatment [6], since professional help involves expensive clinical procedures [7]. The medical cost of mental health exceeds $200 billion in the U.S., making it the costliest medical condition in the country [8].

The digitalization of health information has created opportunities for individuals to seek self-help and connect directly to other affected individuals [9]. Online information

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

is free, anonymous, and time- and location-independent [11]. 80% of the U.S. population with Internet access gather information from mental health discussions, and 34% of those read others' personal stories [12]. Mental health forums are particularly appealing to those individuals who are afraid that coming out of the cloak of anonymity may expose them to stigma [33]. Online information is by no means limited only to the younger generation, as the usage of social media by adults aged 60 and over nearly doubles on a yearly basis [13]. Furthermore, studies show that users are more honest and more likely to share personal stories online than in-person [14]. The chance to share experiences, connect with others with similar conditions as well as gain insights from their stories, creates a rewarding experience for the users of these forums [12].

This paper looks into online tools for mental health by longitudinally analyzing sentiment development of user posts in online mental health forums. We investigate (1) the sentiment progression in forums over time, and (2) the user role dynamics, as well as (3) the relation between user role and sentiment. We apply sentiment analysis on a dataset of 500,754 individual posts across 8 mental health conditions collected from 3 leading English-language forums. We show that the longitudinal development of user sentiment differs across conditions and types of engagement. We compare user roles and their correlation to sentiment.

This study contributes to existing research by exploring how engagement in virtual healthcare communities affects users, and the potential to empower patients to self-management. We especially address the application of behavioral analytics for mental health, which has been more widely adopted for commercial purposes [16], by illustrating how text mining can help practitioners and policy-makers in understanding the value and risks of using user-led online tools [17].

2 Conceptual Background

2.1 Online Mental Health Forums

Previous IS research has focused on trust formation in online health communities, triaging of symptoms, user roles in forums [18, 20], but not on forum influence on well-being. We address this research gap by looking into conversation dynamics and their influence on user sentiment. Prior research shows that users recognize risks of posting personal medical information online, they nevertheless share as potential rewards outweigh risks [19]. User roles have been researched in terms of superusers (frequent posters), however not in terms of sentiment and its correlation to user roles.

Prior research has found online communities to be not only helpful for mitigating various mental conditions but also dangerous, as users can be influenced to commit potentially life-threatening actions, as in pro-suicide or pro-anorexia groups [21]. Many of these studies have applied manual analysis, such as user surveys or discourse analysis based on a small sample of posts. Johnsen, Rosenvinge and Gammon [22] used human readers to classify interactions in mental health forums as helpful or unhelpful based on only 102 posts. By manual assessment, Spijkerman, Pots and Bohlmeijer [23] investigated advantages of online mindfulness and meditation practice for depression and anxiety. In a comparable paper, Mitchell et. al. [24] used human coders to analyze

401 posts from 55 forum threads and found that 25% of users with ADD reported positive effects of self-medicated cannabis on their illness. Although human readers provide reliable analysis, the amount of processed posts is very limited. In this study, we analyze 500,754 forum user posts. Such a vast dataset, to the best of our knowledge, has not been analyzed yet in this context, thus allowing us to explore the shared stories and experiences of tens of thousands of individuals.

2.2 Sentiment Analysis

Recently, natural language processing (NLP) techniques such as sentiment analysis (SA), emotion classification, and stigma measurements have been increasingly applied in research on various data sources such as Twitter, Facebook, and online forums [41]. SA can determine sentiment polarity in written text [25], using a classifier such as Naive Bayes to train with a pre-annotated dataset of sentences, and then apply to new data. Sentiment in psychological literature is referred to as mood or affect. There is a long-standing position in psychological research that a well-adjusted mood is crucial for good mental health [26, 27]. A person in good mental health would express about "three times more positive than negative affect" [28]. Diehl et. al. formally showed that the absence or presence of positivity or negativity can be used to distinguish an individual's mental health status (waning vs. healthy). Therefore, while sentiment is not a tell-all signal of mental well-being, it is a fundamental indicator of the progression of a mental state.

Only few papers use SA in the context of online mental health forums. Nguyen et. al. [29] conducted a fundamental effort by comparing the sentiment expressed on depression forums with sentiment in non-depression forums, demonstrating that individuals without depression express themselves more positively. The study directly backs the application of SA as an appropriate method to analyze data from mental health forums. SA has also been applied to pharmacovigilance in social media (Twitter) – the identification of adverse drug effects – by detecting posts with negative sentiment towards specific medications [30], allowing researchers to explore drug effects across a large and diverse population. The study presents an important example in using text mining to identify potentially negative effects of treatments that might have been otherwise considered safe. Twitter alone has been used in a dozen studies where healthcare issues in general have been explored using sentiment analysis [43]. Furthermore, Cobb, Mays and Graham [31] demonstrated by using SA that talking positively about quitting smoking influences social media users to quit in real life. Via SA, the study shows that user-to-user communication can affect life choices. Aspect-based sentiment analysis has also been applied in order to examine how forum users express themselves on specific concepts such as family or therapy [42]. Finally, Wang et. al. [32] applied SA to classify a user's condition. These ways of using SA have had great success and will be even more useful in the future, considering the development of automated online diagnostic tools [6]. What has not been addressed is the effect of participating in mental health forum conversations, and if and how this depends on the way a user engages with the forum and the role they assume.

3 Research Propositions

Many studies have provided evidence that exchanging support and talking to peers can improve well-being [34]. In fact, even if people seek treatment from professionals, they often still participate in online discussions [33]. Participants also enhance their psychological well-being by providing support [35]. As a result, they feel "better informed, confident with the physician, treatment and social environment, improved acceptance of the illness, increased optimism and control, enhanced self-esteem" [34]. Disadvantageous posts may also occur due to the lack of control of quality of information, potentially destructive content that reinforces negative emotions [10, 22]. Some evidence of the possible benefits from participation exists specifically for mental health forums, although findings are mixed mainly due inconclusive studies and a lack of quantitative research [15].

Owing to the exploratory nature of the study, we use propositions to frame our central premises. In our study we look explicitly at forum users who post, therefore in this context users (or participants) are those who post on the forums, whether to ask a question, to share an experience, to provide answers or comment. By applying sentiment analysis, we test if through participation in forum conversations the affect of a participant is improved. Since threads in mental help forums are started with the purpose to solve or at least mitigate the problem of the original poster [15, 34], the sentiment is expected to improve throughout the thread if the conversation is to be proven as helpful.

P1a: The sentiment score of posts within a thread in mental health forums increases throughout the thread with each subsequent post.

Moreover, we expect that after receiving advice and support, the original poster (OP) will improve their mood, e.g. with regards to confidence, optimism, or ability to cope with their mental condition [15, 34, 36] .

P1b: The sentiment score of posts by the original poster within a thread in mental health forums increases throughout the thread with each subsequent post.

Second, we investigate whether the proposed positive sentiment progression is true outside of single threads, i.e. throughout a user's general forum participation. Studies [36] found that "helping others" was one of the empowering processes that occurred the least frequently in social support groups, since users have to actively post to assist others. High-frequency posters were found to provide more advice than asking for help [36]. Therefore, if a user is commenting on threads by other users, they will improve sentiment and keep the solution-oriented perspective.

P2: The sentiment score of user posts in mental health forums increases throughout the general participation in the forum with each subsequent post.

Third, we look into the roles that users assume in mental health forums, and the relation between roles and sentiment. Initially, users participate by passively reading [37]. In urgent cases, they will post their problem fast and receive encouragement or advice. However, users also increase their psychological well-being by commenting on and engaging with issues posted by others [35], and providing help to others will indirectly assist in coping with their initial problem. Therefore, we propose that users of mental health forums normally fall into one of two roles, either original posters or commenters. The interaction between these two user groups is what drives and defines mental health forum dynamics [38]. Proposition P3 is sentiment-independent and helps establish the dynamics of a user's forum lifetime. The analysis procedure for P3 is to look at the timeline and type of user comments - specifically original poster versus commenter.

P3: Users start their forum participation as original posters, then go on to participate and reply to threads opened by other users.

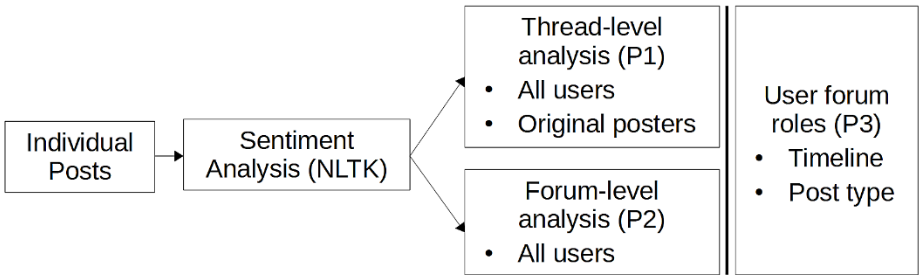


Figure 1. Research Process and Propositions

4 Methodology

To test our propositions, we apply sentiment analysis on peer-to-peer mental health forum posts. Three of the leading English-language mental health forums were scraped to create a combined set of 49,113 threads containing 500,754 individual posts from about 75,000 users across 8 conditions (i.e., depression, bipolar disorder (BD), anxiety and panic attacks, attention deficit hyperactivity disorder (ADHD), Borderline personality disorder (BPD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD)). After removing outlying users who have posted more than 30 times, each user posted on average 6,7 posts throughout a forum. The data were extracted in August 2017 and encompass all publicly available posts on the respective websites. For each post, the dataset contains information on forum username, date posted, thread name, and post content. The forums have moderators whose task is to make sure conversations do not go off-topic; thus, we can be sure that in our research we are considering discussions relevant to each condition. Additionally, moderators remove offensive or damaging material (e.g. posts that encourage self-abusive behavior). However, they do not provide advice, as the forum is a place of discussion among the users, i.e. a forum is not meant or seen as a tool to replace established practices such as therapy.

The sentiment score of a post is scored between -1 (negative) and +1 (positive). We use the Python Natural Language Toolkit (NLTK) sentiment analysis implementation with a lexicon-based classifier. Each forum post is given a score. In order to trace the sentiment progression of user posts throughout forums and threads, we map the average sentiment score of individual users' posts based on the post order within threads (P1) and forums (P2); for proposition P1b we look exclusively at posts made by the original poster within the threads the OP started.

The sentiment is calculated using the Valence-Aware Dictionary for Sentiment Reasoning (Vader) (Gilbert & Hutto, 2014), which offers several unique advantages to other models. The sentiment of a post equals the sentiment valence (or score) of each word recognized by the lexicon. If a word is part of a negation structure (e.g. neither…nor), it's valence will be reversed. If a word is used in combination with a booster word (e.g. "amazingly", "awfully"), its valence will be intensified. The Vader lexicon was built on social media data, namely Twitter tweets, New York Time (NYT) articles, and online movie and product reviews. The sources capture a variety of aspects of social media writing, as well as more analytical texts, thus generalizing well to mental health forum data as it captures features from informal online discussions, e.g. conventional use of punctuation ("good" vs. "good!!!"), capitalization, emoticons, degree modifiers ("good" vs. "very good"), as well as common slang and abbreviations ("this sux", "lol") to signal sentiment intensification. Since its creation in 2014, Vader has been iteratively empirically validated by human judges.

In order to determine the extent of improvement resulting from a prolonged forum posting, we test sentiment score trends by applying regression analysis. For P3, we test the association between user roles and conditions by applying chi-square analysis.

5 Results

Using linear regression, we analyze the sentiment score of a sequence of posts made over a period of time, either within threads or on a per-user basis. Equation (1) models the forum-wide sentiment of depression.

$$y = 0.0011 * x + 0.09 \tag{1}$$

Table 1 presents the regression results and significance for degree of improvement for sentiment trends from three points of view: (1) thread sentiment progression for OP posts only, (2) thread sentiment progression for all thread participants, and (3) sentiment progression per user throughout a forum. Table 1 shows a mild improvement in affect as users keep posting for some conditions, and deterioration in other conditions. The posts of original posters show the highest improvement. The pace of change in sentiment on forums (whether positive or negative) is slightly slower compared to patient improvement observed in therapy. For comparison, a moderate therapy treatment would last 3-4 months and would include 15-20 sessions for 50% of patients to report significant symptom improvement (Leichsenring et. al., 2004). According to our analysis, anxiety, bipolar, and depression forum users must post 100 posts in order to improve their expressed sentiment by 15%. For example, if a depression forum user posts as an OP in threads, the average sentiment improvement per post is 0,003, thus adding +0,3 to positive sentiment after 100 posts. The low R2 values point to high variability in individual user post sentiments. This is consistent with observations from previous studies with a psychological or sociological aspect, as individuals are relatively meandering in their responses [40]. Testing for weekday effects on sentiment change did not prove to be significant.

Table 1. Regression Results and P-Values for Analyzed Mental Conditions

	OP Thread Posts		*All Thread Posts*		*All Forum Posts*	
	Coeff.	**R^2**	**Coeff.**	**R^2**	**Coeff.**	**R^2**
ADHD	0.007	0.2	0.0001	0.2	0	0.2
Anxiety	0.003*	0.3	0.001*	0.4	0.001*	0.3
Autism	0.002*	0.3	-0.006**	0.3	0.002	0.3
Bipolar	0.003**	0.4	0.001	0.2	0.001**	0.2
Borderline	0.001*	0.4	0.0001	0.2	0.0001	0.2
Depression	0.003**	0.8	0.001*	0.4	0.001**	0.4
OCD	0.0001*	0.2	-0.001*	0.2	-0.001	0.1
PTSD	0.0001	0.4	0.003**	0.5	0.001**	0.4

Note: N = 500,754; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;, Coeff = Coefficient

5.1 Thread Sentiment Progression of Posts by all Users

Overall, 5 out of 8 tested conditions (i.e., depression, anxiety, autism, PTSD, OCD) show statistically significant sentiment trend. For depression, anxiety, and PTSD, as the conversation thread progresses, post sentiment improves marginally for every subsequent post. PTSD threads have the steepest improvement per post. On the other hand, autism and OCD mark a downward trend, with autism sentiment deteriorating the fastest per post.

Also, OCD and autism threads show significantly more fluctuation within thread posts, whereas conversations in depression, autism, and PTSD have relatively stable trend progressions.

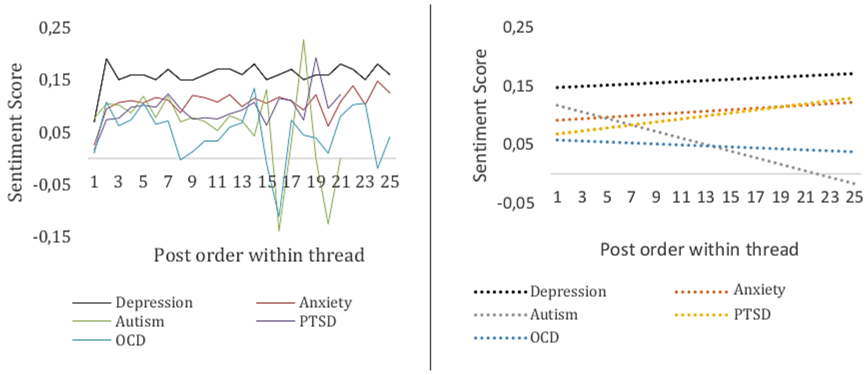


Figure 2. Average Post Sentiment Score Progression within Threads

Based on these findings, Proposition P1a can be confirmed only for depression, anxiety, and PTSD. Thus, mental health forums for these three conditions increase the expressed sentiment in threads that originally started with a lower sentiment score.

5.2 Sentiment Progression of Posts by Original Posters

Across all conditions, OP posts within threads have a positive sentiment score development for every subsequent OP post made (see Fig. 3). However, only 5 out of 8 conditions show significant trends (depression, anxiety, borderline, autism, bipolar). Autism, anxiety, and depression sentiment improve the fastest, whereas depression and borderline forum sentiment improves still, however at a slower pace.

For the most part, the depresson posts by OPs have the highest average score per post order. The anxiety forum exhibits the highest sentiment fluctuation per OP post; interestingly, if an OP opens a thread about anxiety and then immediately follows their own opening post by a second post (usually to provide more detail to their situation), the second post normally has a much lower sentiment score. Thus, those anxiety forum users may be feeling particularly negative when opening a thread, making the subsequent sentiment improvement within their own thread posts that much more substantial.

Based on these results, proposition P1b is supported for 5 out of the 8 conditions tested, showing that, as OPs or advice seekers become more engaged with the thread conversations they start, their expressed sentiment increases.

In terms of improvement per OP post, in all statistically significant conditions, autism forums show the fastest positive sentiment development. This is starkly opposed to general autism posts in a thread, where autism had the worst performance. It is important to note that OCD forums exhibit the same pattern, but without as drastic a change as in the case of autism.

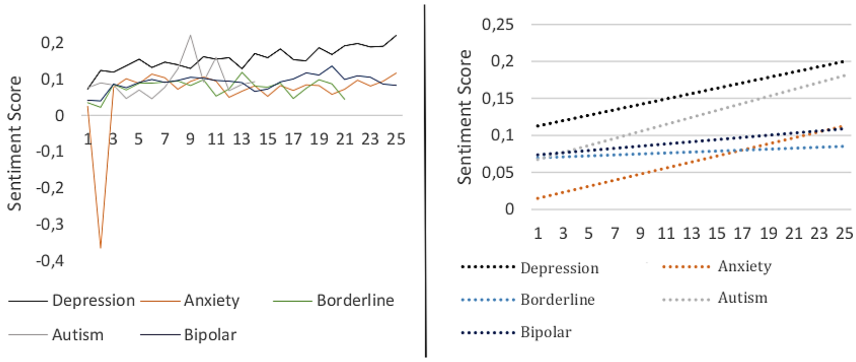


Figure 3. Average Sentiment of OP Post within Threads

Compared to OP-only thread posts, overall thread posts still do improve sentiment score with each subsequent post, however at a slower pace. The coefficients suggest that a forum participant benefits much more from OP posts. Therefore, the forum platforms could encourage users to open their own posts as a way to accelerate and strengthen the increased positivity that results from this interaction.

Additionally, many OPs (50% across all conditions) post only once within the threads they open. However, analysis results indicate that a prolonged participation in the thread which an OP starts actually improves the sentiment within OP posts. Therefore, assuming that the personality of forum members does not play a confounding role, our results suggest that users should be incentivized or encouraged to remain active posters in the threads they create, to draw maximum benefit from their participation.

5.3 Sentiment Progression of Forum-wide Posts by all Users

The 4 out of 8 tested conditions with statistically significant trends are depression, anxiety, PTSD and bipolar disorder, where all show a positive change in sentiment score as a user writes more posts. PTSD marks the highest improvement rates per comment. Of all conditions, only OCD forum users show a negative sentiment score trend with subsequent posts, however the trend is statistically insignificant.

With regards to depression forums, those users with 10 or more posts express a consistently positive sentiment score. Furthermore, the sentiment improves continuously as users post more frequently (regardless of whether as OP or commenter). In the anxiety forum, those users with 42 or more posts express a consistently positive sentiment score. The results point that depression users on average achieve consistently more positive sentiment much faster than anxiety users. PTSD and OCD forum posts show much more variability and fluctuation between positive/negative sentiment. When discussing OP posts vs. thread posts vs. Forum posts (OP and commenter) – the forum-wide posts for the 4 conditions above, even though in general showing positive trend development, mark much more fluctuation in sentiment than the other two post types.

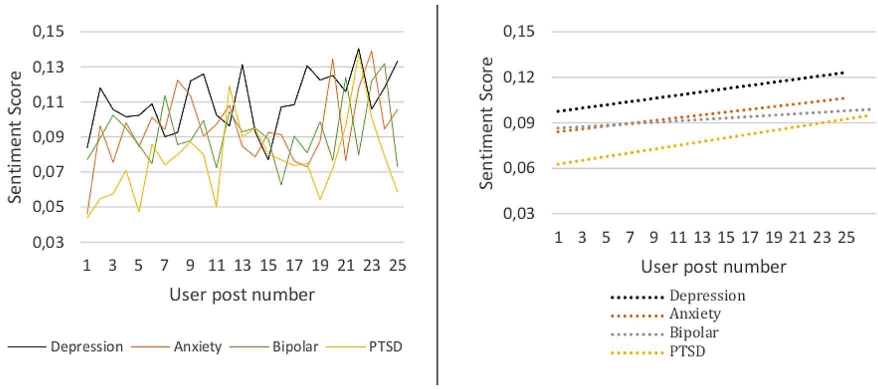


Figure 4. Average Post Sentiment Score per User Post Order within Forums

Based on the results presented, proposition P2 is supported for depression, anxiety, PTSD and bipolar forums, where exposure to mental health forums is shown to be beneficial in terms of improving a user's post sentiment per each subsequent post. For propositions P1 and P2, we can rule out familiarity as a reason for increased sentiment, as the sentiment does not appear to increase or decrease significantly faster depending on how long a user has used a forum.

5.4 User Roles Within Forums

Figure 5 shows the partition of users according to roles for each of the eight conditions examined (chi-square test significance < 0.01). We recognize three user roles: users who exclusively use forums as original posters (OP), users who exclusively act as commenters, and users who take on both roles. A very low percentage of users (15%) take on both roles, which suggests that a transition from OP to commenter or vice versa does not happen for most forum users. PTSD forums have the highest percentage of users who only post as OP (39%), as opposed to bipolar forums, where only 20% of all users are only OP.

Overall, around half of participants across forums use the forums only in a capacity of commenters, with bipolar forums having the highest percentage (64%) of users in that category. The results are concurrent with the reader-to-leader framework [39] in that only few individuals make the transfer from one group to another. Looking into the connection between user roles and sentiment scores, the average OP post sentiment across conditions is always lower than the average sentiment for commenters (see Table 2).

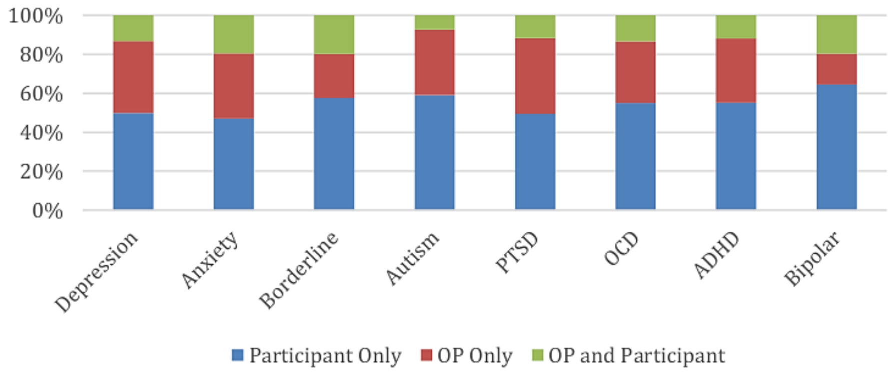


Figure 5. Distribution of User Roles across Conditions

According to the analysis results, there is a correlation between user role and sentiment score. Psychological literature supports the fact that commenters gain benefits by helping other users [36], which is also supported by our analysis. Therefore, forums may be able to further help users in best utilizing forum resources by encouraging participation into the questions and concerns posed by other users.

Table 2. Average Sentiment per User Role across Conditions

	Depres-sion	*Anxie-ty*	*Border-line*	*Autism*	*PTSD*	*OCD*	*ADHD*	*Bipolar*
OP	0.13	0.07	0.07	0.08	0.04	0.03	0.06	0.08
Com-menter	0.16	0.11	0.08	0.10	0.10	0.08	0.09	0.10

6 Discussion and Contributions

Based on our results, in their current form, mental health forums are most beneficial for depression and anxiety sufferers for all roles examined: original poster, commenter, and general forum participant. Some, yet limited benefit was observed for bipolar, borderline, PTSD, OCD, and autism forums, and no benefit for ADHD forums. Borderline forum users benefit only by posting as OP, as opposed to OCD users who only significantly benefit from general thread participation. OCD and autism forum users are the only ones to mark a worsening when participating in threads, specifically in the role of posters (not OP). Research into the content and discourse of the threads and a comparison with contents of other forums can uncover potential conversational and topical differences which might be the reason for opposing sentiment trends. Autism forum users only mark sentiment increase from OP threads. Consequently, autism forum users may benefit from having more options to share their personal stories. These differences in results are important to understand how current forum formats and specific features affect help seekers, and what new features may be beneficial to introduce in order to facilitate forum usage and maximize benefit for all users. The observations made in this study point to a need to provide more customized options for forum participation based on a user's condition – not every type of participation is equally beneficial for all types of conditions and such customizations could optimize a user's forum experience.

In this exploratory study, we apply automated text mining techniques (sentiment analysis) to provide evidence of the benefit of mental health forums. Specifically, we expand the practical application of text mining to mental health behaviors online, thus showing the potential of this technique to not only describe the behavior of thousands of users online, but also to shed light on the environment in which specific users benefit from forums the most. We show that not all users benefit the same, i.e. that mental health conditions as well as user roles are factors related to expressed sentiment. So far, mental health forums have not in any way been optimized or open to customization and personalization, and this paper shows that the individual users stand to gain from such a thing.

We contribute to research by exploring how engagement in virtual user-driven healthcare communities affects users who suffer from mental health conditions, and the potential to empower patients to self-management. Forum creators and administrators can learn from forum content with behavioral analytics in order to adjust forum mechanisms so that they register positive effects across all conditions, and not just a few. At the example of three of the leading English-language mental health forums, we show which forums are beneficial for specific conditions, so that forum creators can guide the future development of these platforms accordingly. We contribute to healthcare analytics research by demonstrating that machine learning and text analytics can uncover new information on user behavior to be used by practitioners and policy-makers in order to advance forum design. We specifically show that forums can in fact have a negative effect on user sentiment for conditions such as autism or OCD, with a further important observation that a change in user roles (e.g. from commenter to original poster) can play a significant role in sentiment development over time.

7 Limitations and Directions for Future Research

The conducted study presents an initial investigation and, thus, needs to be understood with respect to limitations. These limitations simultaneously represent opportunities for future research.

Our study does not control for all factors internal and external to mental health forums, e.g. moderation and feature effects. Also, we did not control whether users received therapy during participation in the forums, which may have affected forum sentiment. Furthermore, using different measures can help form a more detailed picture of forum effects across conditions. Given the difference in positive vs. negative sentiment development in different conditions and for different user roles, future studies can look into the specific forum mood drivers and mechanisms on a per-condition basis. Each of the investigated illnesses are independently unique, with nuances that cannot be fully captured by sentiment analysis alone; the application of various measures and an inquiry specifically regarding forum design and moderation may help to provide more detailed answers to this question. Adopting a user perspective, future studies can also address sentiment as a function of user forum life. Finally, this study presents an application of sentiment analysis on a large dataset with sensitive personal data. This raises privacy concerns which require a detailed separate study.

The digitization of health information has created opportunities for individuals to take control of their health, highlighting the evolving socio-technical change that occurs within healthcare [9]. Although our findings highlight the potential applicability of machine learning within mental healthcare practice and research, our analyses are still an initial endeavor in form of an exploratory approach. Further investigation is needed to understand how and why sentiment develop the way they do, and might be a helpful undertaking to comprehend why health forums are visited so often for advice.

References

1. WHO, http://www.who.int/mediacentre/factsheets/fs396/en (Accessed: 11.06.2018)
2. WHO, http://www.who.int/news-room/fact-sheets/detail/mental-health-of-olderadults (Accessed: 11.06.2018)
3. WHO, http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ (Accessed: 11.06.2018)
4. Mathers, C., Boerma, T., Fat, D.M.: The global burden of disease: 2004 update: World health organisation. (2008)
5. Crabtree, J.W., Haslam, S.A., Postmes, T., Haslam, C.: Mental health support groups, stigma, and self-esteem: positive and negative implications of group identification. Journal of Social Issues 66, 553-569 (2010)
6. Saleem, S., Pacula, M., Chasin, R., Kumar, R., Prasad, R., Crystal, M., Marx, B., Sloan, D., Vasterling, J., Speroff, T.: Automatic detection of psychological distress indicators in online forum posts. In: Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific, pp. 1-4. IEEE, (2012)
7. Weathers, F.W., Keane, T.M., Davidson, J.R.: Clinician-Administered PTSD Scale: A review of the first ten years of research. Depression and anxiety 13, 132-156 (2001)
8. Larner, L., http://www.healthaffairs.org/do/10.1377/hblog20160518.054918/full/ (Accessed: 11.07.2018)
9. Lluch, M.: Healthcare professionals' organisational barriers to health information technologies—A literature review. International journal of medical informatics 80, 849-862 (2011)
10. Finn, J.: An exploration of helping processes in an online self-help group focusing on issues of disability. Health & Social Work 24, 220-231 (1999)
11. Winzelberg, A.: The analysis of an electronic support group for individuals with eating disorders. Computers in Human Behavior 13, 393-407 (1997)
12. Malmasi, S., Zampieri, M., Dras, M.: Predicting post severity in mental health forums. In: Proceedings of the third workshop on computational lingusitics and clinical psychology, pp. 133-137. (2016)
13. Madden, Mary. "Older adults and social media: Social networking use among those ages 50 and older nearly doubled over the past year." Pew Internet & American Life Project (2010).
14. Barak, A., Gluck-Ofri, O.: Degree and reciprocity of self-disclosure in online forums. CyberPsychology & Behavior 10, 407-417 (2007)

15. Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., Stern, A.: Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. Bmj 328, 1166 (2004)
16. Monteith, S., Glenn, T., Geddes, J., Bauer, M.: Big data are coming to psychiatry: a general introduction. International journal of bipolar disorders 3, 21 (2015)
17. Smailhodzic, E., Boonstra, A., Langley, D.: Social media disruptive change in healthcare: Responses of healthcare providers? In: ECIS, pp. Research-in-Progress Paper17. (2016)
18. Huh, J., Pratt, W.: Weaving clinical expertise in online health communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1355-1364. ACM, (2014)
19. Kordzadeh, N., Warren, J.: Communicating Personal Health Information in Virtual Health Communities: An Integration of Privacy Calculus Model and Affective Commitment. Journal of the Association for Information Systems 18, (2017)
20. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. ICWSM 13, 1-10 (2013)
21. Bell, V.: Online information, extreme communities and internet therapy: Is the internet good for our mental health? Journal of mental health 16, 445-457 (2007)
22. Johnsen, J.A.K., Rosenvinge, J.H., Gammon, D.: Online group interaction and mental health: An analysis of three online discussion forums. Scandinavian Journal of Psychology 43, 445-449 (2002)
23. Spijkerman, M., Pots, W.T.M., Bohlmeijer, E.T.: Effectiveness of online mindfulness-based interventions in improving mental health: A review and metaanalysis of randomised controlled trials. Clinical psychology review 45, 102-114 (2016)
24. Mitchell, J.T., Sweitzer, M.M., Tunno, A.M., Kollins, S.H., McClernon, F.J.: "I use weed for my ADHD": a qualitative analysis of online forum discussions on cannabis use and ADHD. PloS one 11, (2016)
25. Tao, Y.: Sentiment Analysis Method Review in Information Systems Research. (2014)
26. Bradburn, N.M.: The structure of psychological well-being. (1969)
27. Kahneman, D., Diener, E., Schwarz, N.: Well-being: Foundations of hedonic psychology. Russell Sage Foundation (1999)
28. Diehl, M., Hay, E.L., Berg, K.M.: The ratio between positive and negative affect and flourishing mental health across adulthood. Aging & Mental Health 15, 882-893 (2011)
29. Nguyen, T., Phung, D., Dao, B., Venkatesh, S., Berk, M.: Affective and content analysis of online depression communities. IEEE Transactions on Affective Computing 5, 217-226 (2014)
30. Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., Gonzalez, G.H.: Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. Journal of biomedical informatics 62, 148-158 (2016)
31. Cobb, N.K., Mays, D., Graham, A.L.: Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. Journal of the National Cancer Institute Monographs 2013, 224-230 (2013)
32. Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., Bao, Z.: A depression detection model based on sentiment analysis in micro-blog social network. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 201-213. Springer, (2013)
33. Kummervold, P.E., Gammon, D., Bergvik, S., Johnsen, J.-A.K., Hasvold, T., Rosenvinge, J.H.: Social support in a wired world: use of online mental health forums in Norway. Nordic journal of psychiatry 56, 59-65 (2002)
34. van Uden-Kraan, C.F., Drossaert, C.H., Taal, E., Shaw, B.R., Seydel, E.R., van de Laar, M.A.: Empowering processes and outcomes of participation in online support groups for

patients with breast cancer, arthritis, or fibromyalgia. Qualitative health research 18, 405-417 (2008)
35. Riessman, F.: Ten self-help principles. Social Policy 27, 6-12 (1997)
36. van Uden-Kraan, C.F., Drossaert, C.H., Taal, E., Seydel, E.R., van de Laar, M.A.: Participation in online patient support groups endorses patients' empowerment. Patient education and counseling 74, 61-69 (2009)
37. Preece, J., Nonnecke, B., Andrews, D.: The top five reasons for lurking: improving community experiences for everyone. Computers in human behavior 20, 201-223 (2004)
38. Becker, L.C.: Reciprocity (Routledge Revivals). Routledge (2014)
39. Preece, J., Shneiderman, B.: The reader-to-leader framework: Motivating technology-mediated social participation. AIS transactions on human-computer interaction 1, 5 (2009)
40. Bedeian, A., Mossholder, K.: Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. Journal of Management 20, 1, 159-165 (1994)
41. Calvo, R. A., Milne, D. N., Hussain, M. S., Christensen, H.: Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering 23, 5, 649-685 (2017)
42. Davcheva, E.: Text Mining Mental Health Forums - Learning from User Experiences. European Conference on Information Systems (2018)
43. Gohil, S., Vuik, S., Darzi, A.: Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health and Surveillance 4, 2 (2018)

Intent and the Use of Wearables in the Workplace – A Model Development

Severin Oesterle[1], Bianca Trübenbach[2], and Christoph Buck[2]

[1] FIM Research Center, University of Bayreuth, Bayreuth, Germany
severin.oesterle@fim-rc.de
[2] Chair of Information Systems Management, University of Bayreuth, Bayreuth, Germany
{bianca.trübenbach,christoph.buck}@uni-bayreuth.de

Abstract. Due to reasons like demographic changes and variations in the spectrum of illness, worldwide expenditures in the health market have exploded. Contemporary information systems are evolving rapidly in the field of ubiquitous computing and nowadays support health in various fields. Wearables and tracking technologies have emerged in private life for health and fitness support. This adoption reveals future possibilities for innovating the health-supporting systems in the workplace. The crucial point of introducing wearables in the occupational health management system is the acceptance of employees. This paper provides a literature-driven measurement model to explain the behavioral intention to use wearables in the occupational health management system. The model provided is supported by 17 hypothesized relationships between relevant constructs and validated by card-sorting.

Keywords: Wearables Devices, Occupational Health Management, Technology Acceptance

1 Introduction and Motivation

Demographic developments are changing the labor market and employee demands on the working world. The world population is ageing, and the proportion of young skilled workers in developed countries is shrinking [1]. As a result of demographic change and changes in the spectrum of illnesses towards chronic degenerative diseases such as diabetes, degenerative musculoskeletal disorders, psychologically manifested diseases and addictions [2], health expenditure is at a record level [3]. In addition, a change in the value system of young employees toward more individuality and work-life balance can be observed [1]. The progressive digitalization requires an increased work speed, greater flexibility, and above all permanent willingness to learn and perform on the part of employees [1]. Not only since increases in digitalization has the workplace has been identified as the leading cause of many mental or psychological illnesses, as well as the primary cause of stress [4]. Various burdens on employees, as well as the developments mentioned above, entail an increased risk of long absences of employees in companies [5]. Workers with poor physical or mental health are often less productive, make worse

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

decisions, and have more absenteeism overall [6]. The employee absenteeism development in Germany illustrates this problem, which is of great economic relevance.

One way of reducing sick leave and creating additional incentives for employees in times of skilled-worker shortage is the introduction of an innovative occupational health management system (OHM). Both science and practice have recognized the significant positive influence of OHM on employees and organizations [7, 8].

At the same time, wearables are becoming increasingly popular with the majority of employees [9]. Research on intelligent portable systems has therefore increased in the health sector under the headings electronic health (eHealth) and mobile health (mHealth), as well as in industry [10]. This trend towards wearables can be seen in the high number of wearable manufacturers and strong industry growth.

Meanwhile, some companies are offering their employees the opportunity to participate in digital health programs in the workplace while following their health and fitness activities through wearables [11]. Studies have already shown that wearables can improve the health and well-being of individuals: Participants in health programs are happy to monitor, track and review their health data and the control of personal health data also encourages participants to behave healthier [12]. Similarly, it was found that the tracking and observing of one's physical activity, for example, leads to an increase in steps taken [13]. Furthermore, health programs which use portable devices increase the average employee participation from 20% to 60-70% [14].

Mainly due to their unique characteristics, wearables are very suitable for supporting the OHM of the future. Wearables can be worn on the body and can be used freely at work. In addition, the devices are mobile, always active and often context-sensitive. Furthermore, employees can use wearables in all areas of the company so that an organization-wide OHM is possible. Also, the possible use of gamification strategies [15] to increase employee motivation and the possibilities for personalizing health care through wearables are also promising.

However, from the employees' perspective, there are serious barriers which prevent participation in company health promotion measures. Many employees do not recognize the benefits of the measures or perceive OHM as paternalism and are afraid to embarrass themselves. Notably, the use of wearables which collect health data could be perceived as a high risk by employees. The fear of health data being misused (e.g., due to inadmissible performance monitoring of employees, as justification of salary increases, promotions or dismissals) based on the data collected is high and requires trust in the employer.

Against this background, the relevance of employee acceptance for new technologies at OHM becomes clear. Technologies and wearables cannot help to improve individual health if employees do not accept and use the technology provided. Some studies are already investigating the acceptance of wearables [3, 16, 17] and other portable technologies in healthcare [4, 18, 19]. Whereas the previous studies serve as valuable starting points, some specific aspects in the use of wearables in the working environment are missing or measure individual consumer's acceptance. Knowing the employee's level of acceptance for specific measures is necessary to increase the success of OHM. Companies must be able to predict employees' acceptance of wearables. However, as far as we know, there has been no empirical research on the acceptance of using wearables for health promotion in the workplace.

Based on a systematic literature review, we set out to deductively develop a research model that will help us to gain a better understanding of wearable acceptance in the workplace. The remainder of this paper is structured as follows: The next section describes the underlying theoretical foundations based on previous literature. The succeeding section introduces the research process, which is followed by the structural model development section. Subsequently, we provide insights into the measurement model development before we conclude with our discussion and outlook.

2 Relevant Work

2.1 Wearable Technologies in Healthcare

Wearables are specific technologies of the Internet of Things (IoT) and a concretization of ubiquitous computing aimed at improving our environment through visible or invisible, networked and intelligent electronic devices which enable a new form of human-computer interaction [20]. There are several different kinds of wearable definitions. However, for our purpose, we define wearable devices as electronic devices that can be worn on the body and measure via one or more sensors physical activities or health conditions and are (wirelessly) connected to other computers or smart devices [21]. The sensors collect a broad range of different data, such as various vital signs, physiological parameters, and environmental conditions [22, 23].

In addition to the use of wearables in health science, there are also many other fields of application, e.g., consumer goods [24], in the fields of professional and recreational sports [25], and for authorities with security tasks [26]. Within the application of wearables in the health context, a large number of overlapping terms are used. The most common definition defines eHealth as "an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve healthcare locally, regionally, and worldwide by using information and communication technology" [27, p.1]. Although there are differences between the terms, it is accepted in research to refer to eHealth as an all-encompassing term for telemedicine and mHealth and includes various forms of HIT [10].

The concepts of workplace health prevention and health promotion are complementary and overlap in many respects and therefore are partly used synonymously in practice. We refer to OHM which includes both concepts. A trend towards digitalization is discernible, not only in the health sector, but also in the area of OHM [28]. This digitalized health promotion is a possible field of application for digital health technologies. Modern technologies open up new opportunities for OHM in the areas of requirement analysis and employee awareness.

While OHM causes an increase in the health affinity of employees, promotes high participation rates, and can reach all employee target groups, criticism exists with regard to data privacy [4, 29]. One of the biggest challenges in OHM in connection with wearables relates to the solution of IT security issues, data protection concerns, and legal and ethical difficulties in handling the data. In most cases, wearables forward data to the manufacturer, third-party providers and intermediaries (e.g., insurance companies, scientists, advertising companies, and, in the case of the present work, the employer). The gathered data is often stored decentrally in a cloud. Hence, employers need to guarantee system security and assure data privacy.

In summary, the voluntary and private use of wearables for individuals who are interested in improving their health and fitness is not new. However, the wearables' usage during work and the integration in an OHM become more and more important [30]. The employees' acceptance is essential for an effective wearable usage in OHM. Empirical studies which examine the employee acceptance of OHM with wearables are surprisingly scarce, which is why we set out to deductively develop a concise structural equation model in the following sections.

2.2 Health Behavior

Regarding the field of health behavior, four main theories are used: (1) the Health Belief Model, (2) the Protection Motivation Theory, (3) Subjective Expected Utility Theory and Theory of Reasoned Action [31]. Each of the four theories assumes that an expected negative health event and the desire to avoid or reduce this occurrence will motivate self-protection. Furthermore, they explain health-related behavior on the basis of the expected value theory and the cost-benefit analysis. The Health Belief Model assumes that a person who decides to perform a health-related action takes the action from a consideration of the perceived health threat that occurs when the action is omitted and the belief in the effectiveness of the proposed measure, minus perceived barriers [32].

Moreover, the Health Belief Model is used to develop health promotion measures [33]. It incorporates four basic constructs: perceived susceptibility, perceived severity, perceived benefits and costs, and perceived barriers. The Protection Motivation Theory uses similar factors to explain health behavior (perceived vulnerability, perceived severity, response efficiency, response costs) [34] and in addition, the determinant self-efficacy is integrated into the Protection Motivation Theory. The Health Belief Model's perceived benefits and Protection Motivation Theory's response efficacy measure the same underlying construct and can be equated with the performance expectancy construct of Unified Theory of Acceptance and Use of Technology (UTAUT) [35]. It is argued that the Health Belief Model's perceived benefits are equivalent to the two determinants intrinsic and extrinsic motivation of UTAUT2 [4]. Further determinants also determine preventive health behavior. The Health Belief Model's supplementary determinants include, as preceding factors, e.g. the demographics of the individual, psychological factors, the mediator variable cues to action, advice on health, a letter from the doctor, or an emerging illness within the family or circle of friends, as well as the mediator variables self-efficacy, response efficacy and the value of action [36].

Furthermore, preventive health behavior is influenced by response efficacy and self-efficacy, as well as health motivation and health consciousness [36]. Health motivation "refers to consumers' goal-directed arousal to engage in preventive health behaviors"

[36, p.9], while health consciousness "refers to the degree to which health concerns are integrated into a person's daily activities" [36, p.10].

2.3 Technology Acceptance

Many research disciplines are investigating the adoption or acceptance of innovations. Studies on the acceptance of technologies are regarded as the most important research field in information systems (IS).

In the field of economics, particularly in connection with innovation management, adoption is understood as the acceptance of an innovation by an individual customer. A positive decision to accept an innovation by users is therefore acceptance. There are different views in literature as to whether the attitude towards innovations, the intention of behavior or benefit, the behavior, or a combination of these factors should be regarded as acceptance. In economic acceptance research, the distinction between acceptance of attitudes and acceptance of behavior has, therefore, become established. The intention to use a technology is equated with attitudinal acceptance. Therefore, this type of acceptance is not directly observable by users and is, therefore, operationalized by the behavioral intention. In contrast, when adopting innovations in the form of observable behavior (e.g., the use of a wearable for OHM), acceptance of behavior is discussed. Acceptance is then determined by usage behavior. The wearable technology in the OHM context can be seen as (technical) innovations and can either be accepted or rejected by users, in our case employees. The acceptance of attitudes is operationalized as an intention to use or behave (behavioral intention). The role of behavioral intention as a predictor of behavior has been extensively researched in IS literature and related research fields [37–39]. Behavioral intention is defined as a measure of the strength of an individual's intention to conduct a certain behavior [40].

The rich literature on adoption of technologies, and in particular the proposed Technology Acceptance Model (TAM) of Davis [41], is being studied in detail by IS researchers. However, TAM is also criticized for being too parsimonious to reliably explain complex psychological processes such as behavior and human’s technology acceptance, and it does not take influences of social and personal control factors into account [38, 42]. The results of our literature review indicate that within the plethora of acceptance models there is only one study which provides a model for measuring acceptance of mHealth applications in the OHM context [43]. The study discusses the suitability of mHealth apps for the use in the OHM context and the underlying determinants that motivate employees to use health apps at the workplace. Based on TAM, the Health Belief Model, and the Theory of Planned Behavior, the authors propose a model to explain adoption behavior. However, it does not regard aspects of motivational theory and is quite complex, due to the inclusion of fifteen explanatory constructs. Furthermore, there is no empirical validation of the proposed model. Based on these findings, we take both the specific characteristics and the use of wearables in the context of OHM into account when designing our examination model. Therefore, we transfer existing studies of relevant other contexts and present an overview of acceptance research studies which are germane for our model development.

Since wearables are small body-worn computers and part of the IoT, studies on acceptance research in both areas, as well as studies on different types of wearables, are relevant for our model development. Many researchers have demonstrated the influence of intrinsic and extrinsic motivation, which are often operationalized as perceived enjoyment and perceived usefulness [44, 45].

In addition, a relevant study on the acceptance of IoT [29] and a study on ubiquitous, pervasive technologies [46] were identified. The acceptance of IoT measures the influence of perceived usefulness, perceived ease of use, social influence, and perceived behavioral control, which are equated in the study with the constructs of UTAUT, on behavioral intention. The authors supplement their model with the constructs trust and perceived enjoyment. The results confirm a strong influence of the UTAUT determinants, but a non-significant influence of trust on behavioral intention [29]. The study of ubiquitous technologies also confirm the strong influence of perceived usefulness and perceived ease of use on the behavioral intention of pervasive computers, but cannot confirm a significant influence of compatibility, perceived overall risk, and attractiveness of alternatives [46].

We also identified some studies on wearables. It is striking that most of the studies use different combinations of the UTAUT's main determinants. In addition, the perceived enjoyment construct and intrinsic motivation are often included in the models [3, 17, 47]. The determinants trust toward the employer [17] and other forms of trust [3], as well as different forms of risk [47], are also included in the models for predicting the behavioral intention of wearables. Summarizing, we observe a significant influence of the UTAUT determinants perceived usefulness, social influence, and perceived enjoyment. Within the acceptance research of eHealth technologies, there are also studies on the acceptance of wearable-related technologies [5, 48], such as mHealth technologies [35, 49, 50], Health Information Technology [18, 51], and Wearable Health Monitoring Systems [4]. In contrast, studies on the adoption behavior of health technologies for patients or consumers, which are essential to the requirements for an acceptance study of employees, are relatively rare [4, 50, 52]. Among these empirical studies on users' eHealth adoption behavior, most studies explain the usage behavior based on TAM [18, 50] and its extension UTAUT [4, 35]. The direct and indirect influence of perceived usefulness and perceived ease of use is confirmed in all empirically tested studies. Similar to the acceptance studies for wearables, perceived enjoyment and hedonic motivation were additionally included as determinants [4].

Whereas the identified studies and models serve as a valuable starting point, none of the presented studies focus explicitly on wearables in the OHM context and therefore take the specific requirements of the acceptance of wearables in the OHM into account.

3 Structural Model Development

Based on the scientific literature presented above, we derived our hypotheses and built our structural model. We combined the Health Belief Model, and the Motivation Model, and incorporated factors from UTAUT. We present our developed structural model in Figure 1.

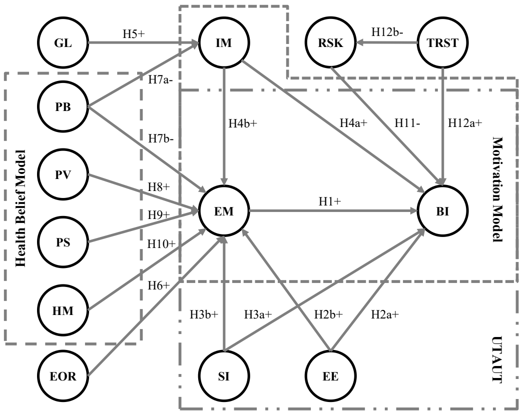


Figure 1. Structural model explaining the behavioral intention

With our model, we set out to explain our dependent variable *behavioral intention to use wearables in OHM* which can be seen as a predictor for acceptance. UTAUT takes the determinant extrinsic motivation of the motivation model as performance expectancy into account [39]. Extrinsic motivation is defined as "a construct that pertains whenever an activity is done in order to attain some separable results" [53, p.60]. Concerning wearables in OHM, extrinsic motivation is defined as the perceived probability that a wearable supports the employee in achieving its goals. Thus, we assume that the prospect of achieving these goals has a positive influence on the behavioral intention which is why we propose:

H1: Extrinsic motivation (EM) has a positive impact on the behavioral intention to use a wearable in OHM (BI).

The determinant effort expectancy [4] or the closely related construct perceived ease of use [29] was also identified as an essential factor for measuring acceptance. Effort expectancy is defined as “the degree of ease associated with the use of the system” [39, p.450]. In addition to the effort expectancy, social influence is also included in suitable acceptance models. The social influence is defined as the "extent to which consumers perceive that important others (e.g., family and friends) believe they should use a particular technology" [54, p.159] and hence, has an impact on behavioral intention and extrinsic motivation. We state that a wearable in OHM is a portable device for increasing and promoting health. We argue that friends and family have a particular interest in the individual employee's health. Thus, the close social environment motivates the individual to take part in OHM. The perceived social pressure and the opinion of an individual's environment have an impact on behavioral intention and extrinsic motivation. Thus, we conclude:

H2a: Effort expectancy (EE) has a positive impact on behavioral intention (BI).
H2b: Effort expectancy (EE) has a positive impact on extrinsic motivation (EM).
H3a: Social influence (SI) has a positive impact on behavioral intention (BI).
H3b: Social influence (SI) has a positive impact on extrinsic motivation (EM).

Intrinsic motivation is defined as "the doing of an activity for its inherent satisfactions rather than for some separable consequence" [53, p.56]. Accordingly, it represents the extent to which employees perceive wearables in OHM, apart from other consequences, as pleasant and entertaining. Studies on consumer behavior as well as research in the field of IS have found that intrinsic motivation is an essential determinant for explaining technology acceptance [54, 55]. Furthermore, in the context of mHealth services and also in the context of wearables, it is shown that the user's intention to use mHealth services is determined by both extrinsic and intrinsic motivation [3, 29, 47, 49]. Thus, if an employee has intrinsic motivation to improve his or her health, it is all the more likely that he or she will participate in OHM activities and take greater account of the benefits provided [56]. We include intrinsic motivation in our model and state that an activity perceived as pleasant or entertaining has a positive influence on the perception of the usefulness and thus, supports a user to achieve its goals. Thus we hypothesize:

H4a: Intrinsic motivation (IM) has a positive impact on behavioral intention (BI).

H4b: Intrinsic motivation (IM) has a positive impact on extrinsic motivation(EM).

When introducing a wearable in OHM, it is crucial that the use of the device is accepted by employees as quickly as possible. So it is vital to convince employees to adopt technological innovations at an early stage. We consider employees with a high motivation to use different technologically innovative devices as gadget lovers. A gadget lover is defined as "a consumer with high intrinsic motivation to adopt and use a variety of leading-edge, technology-based goods, including the services that complement them" [57, p.330]. So far, the gadget-loving concept has yet not been integrated into acceptance research. We incorporate gadget loving as an external variable which measures the employee's personal characteristics. Hence, we postulate:

H5: Gadget loving (GL) has a positive impact on intrinsic motivation (IM).

From the perspective of the extrinsic motivation, a behavior is carried out on the basis of the expected benefit of the action or the expected advantages that an action brings [53]. Expected organizational rewards can be regarded as fundamental objectives of extrinsically motivated behavior [58]. Rewards are also often used in the OHM context to motivate employees to participate in specific measures [6]. These expected organizational rewards can range from monetary incentives, such as discounted membership in gyms, to points in competition with other employees. Thus, the offered rewards shall motivate employees' participation. Hence, we conclude:

H6: Expected organizational rewards (EOR) have a positive impact on extrinsic motivation (EM).

Furthermore, we also included the Health Belief Model in our model. It proposes that perceived barriers are subtracted from the perceived benefits [32]. The negative aspects of a health-promoting activity can act as barriers to the implementation of the recommended behavior. Consequently, if employees consider OHM activities as, for example, unpleasant or time-consuming, their motivation will decrease. On the basis of a cost-benefit analysis, the advantages of preventive measures are weighed against their disadvantages. Since the perceived benefits of the preventive measure are already taken into account by the constructs intrinsic and extrinsic motivation [4, 35], we only include the perceived barriers into our model. Thus, we hypothesize:

H7a: Perceived barriers (PB) have a negative impact on extrinsic motivation (EM).

H7b: Perceived barriers (PB) have a negative impact on intrinsic motivation (IM).

Furthermore, we integrated the effects of the perceived health threat, operationalized as perceived severity and perceived vulnerability, into our model. Perceived vulnerability is defined as the perception of an individual's vulnerability to health threats, while perceived severity is defined as the assessment of an individual as to whether a particular health threat is severe or dangerous [59]. From the perspective of extrinsic motivation, the expected benefit of an action determines the behavior. Employees are expected to value the benefits of the use of wearables in OHM more if the expected consequences of a resulting health threat are serious. Hence, we state:
H8: Perceived vulnerability (PV) of an individual has a positive impact on extrinsic motivation (EM).
H9: Perceived severity (PS) of an individual has a positive impact on extrinsic motivation (EM).

We also incorporate health motivation into our model, which "refers to consumers' goal-directed arousal to engage in preventive health behaviors" [36, p.9] and hence, refers to the internal characteristics. Studies have found that this can be associated with most health behaviors [36]. This preventive health behavior is of great importance to many individuals and represents an essential general goal in life. Hence we conclude:
H10: Health motivation (HM) has a positive impact on extrinsic motivation (EM).

According to the literature on wearables, risk has a significant influence on the behavioral intention of individuals [4, 47]. Within the acceptance research of health-related IT, the fear of unintentional secondary data use and privacy concerns are often investigated [4, 29, 49, 60]. For example, in the case of restructuring measures, employees face the risk that the employer will incorporate the gathered health data in its redundancy decision. Therefore, we define risk as the total perceived threats that employees feel when using wearables in OHM. Hence, we postulated that risk has a negative influence on behavioral intention. Thus, we conclude:
H11: The perceived risk (RSK) has a negative impact on behavioral intention (BI).

The long-term exchange of personal information (e.g., with the employer) through the use of wearables in OHM also requires a trustful relationship between the employee and its employer. In the context of eHealth technologies and also in technology acceptance studies, it is confirmed that the effect of trust is a decisive factor [60, 61]. We argue that employees, in particular, are often in a weaker position than their employers. Especially in our context, highly personalized user data is collected, and there is the chance that the employer may misuse this data. We define trust as the perception of the employee that its employer is trustworthy and wants to do something good for its employees by introducing wearables in the context of OHM. Since trust has a significant influence on the behavioral intention of individuals we thus conclude:
H12a: Trust in the employer (TRST) has a positive impact on behavioral intention (BI).
H12b: Trust in the employer (TRST) has a negative influence on perceived risk (RSK).

4 Measurement Instrument Development

The development of a suitable measurement model is crucial for the causal model's future evaluation. The development of our measurement instrument is therefore carried out in several stages [62]: (1) We identified existing measurement scales where possible and initially created new ones where necessary. The measurement scales for the latent variables are expressed by several manifested statements (items) [63]. (2) Subsequently, we adapted the identified items to our context and gained an initial item long-list. (3) Subsequently, three scientists iteratively reviewed the initial item pool and adjusted the items to fit our context and provide a common style in language and wording. (4) Following this, we conducted a card-sorting procedure proposed by Moore and Benbasat [62]. The card-sorting procedure aims to assess the construct validity of the various scales and to "attempt to identify any particular items which still may have been ambiguous" [62, p.199]. Therefore, we sent out an Excel-spreadsheet containing a VBA macro for randomization via email and asked twelve judges to sort the randomized initial item pool to the corresponding construct with given construct definitions [62, 64]. The group of judges consisted of employees in order to ensure that the items to be evaluated were understandable for future participants. In a second step, after the judges had assigned the items to the corresponding constructs, we asked them to sort the items of each construct according to their representativeness, to then identify the most appropriate items of the initial long-list, and then to send back the spreadsheet. After having received the completed card-sorting, we consolidated the filled-out spreadsheets and analyzed the results. This enables the identification of items which are not suitable to measure the underlying construct. It is assumed to exclude these items from the study [64]. After having received the results, we evaluated the construct validity of our initial item pool and removed items which were mainly sorted into an incorrect construct. (5) Finally, we pilot-tested the resulting item pool with a seven-point Likert-scale and analyzed the gathered feedback to complete our instrument. Thus, we ended up with a total of 51 items for our final measurement model. Table 1 provides the sources and number of items for the final measurement model.

Table 1. Sources of the measurement model

Construct	# of items	Adapted from
Behavioral Intention	4	[41, 65]
Extrinsic Motivation	4	[41, 66]
Effort Expectancy	4	[41, 65]
Social Influence	4	[39]
Intrinsic Motivation	4	[65, 66]
Gadget Loving	4	[57]
Expected Organizational Rewards	4	[67], own
Health Motivation	4	[68, 69]
Perceived Barriers	5	[70, 71], own
Perceived Severity	3	[70]
Perceived Vulnerability	4	[70]
General Risk Beliefs	3	[72]
Employees Trust Beliefs	4	[1]
Total	51	

5 Conclusion and Further Research

In this article, we set out to develop a literature-based measurement model to explain the behavioral intention to use wearables in OHM. As highlighted in the first section, due to demographic developments as well as changes in the spectrum of illness and in the value system of employees, health expenditures in industrialized countries like Germany have exploded. The introduction of modern, easy to use and consumer-centric IS in the workplace possibly meet these challenges. IS, such as wearables, do have the functionality to support employees' health conditions. Consequently, the crucial point of introducing wearables in the workplace for OHM purposes is the employees' behavioral intention to use the provided technologies.

As the presented work shows, many studies exist on the acceptance of several kinds of technologies. Although the possibilities of supporting the health of users by IS are highly relevant, research is surprisingly scarce. Therefore, we reviewed the current state-of-the-art of wearable technologies, health behavior, and technology acceptance from the perspective of possible applications in the workplace. Subsequently, we developed a measurement model to explain the behavioral intention to use wearables in an OHM context. The measurement model was validated by a card-sorting procedure and led to a measurement model consisting of a total of 51 items. With our model, we provide a research tool to explain the behavioral intention to use wearables in OHM. Regarding our specific next steps in this research endeavor, we deem quantitative-empirical methods as most applicable to validate our model. Therefore, we will collect data by conducting a large-scale multinational online survey study and analyze the gathered data using a structural equation model approach [73] and will subsequently strengthen our statistical analysis by conducting multi-group comparisons [74]. The survey will incorporate different kinds of wearables such as wristbands and smart clothing. Hence, differences in wearable technologies in the behavioral intention to use might be identified, and practical implications can be derived. Furthermore, we will apply the research model in different industries and different organizations to gain a deeper insight into under which conditions employees accept the implementation of wearables for OHM. To get a more comprehensive view on the intention to use wearables in the workplace, different theories (e.g., system theory) from related domains could be taken into account as OHM affects a lot of different perspectives, habitats, actors, and systems.

References

1. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce. Information Systems Research 13, 334–359 (2002)
2. Pai, F.-Y., Huang, K.-I.: Applying the Technology Acceptance Model to the introduction of healthcare information systems. Technological Forecasting and Social Change 78, 650–660 (2011)
3. Rupp, M.A., Michaelis, J.R., McConnell, D.S., Smither, J.A.: The Impact of Technological Trust and Self-Determined Motivation on Intentions to use Wearable Fitness Technology.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting 60, 1434–1438 (2016)
4. Gao, Y., Li, H., Luo, Y.: An empirical study of wearable technology acceptance in healthcare. Industrial Management & Data Systems 115, 1704–1723 (2015)
5. Hennemann, S., Beutel, M.E., Zwerenz, R.: Ready for eHealth? Health Professionals' Acceptance and Adoption of eHealth Interventions in Inpatient Routine Care. Journal of Health Communication 22, 274–284 (2017)
6. Albu, M., Atak, L., Srivastava, I.: Simulation and gaming to promote health education: results of a usability test. Health Education Journal 74, 244–254 (2015)
7. Danna, K., Griffin, R.W.: Health and well-being in the workplace: A review and synthesis of the literature. Journal of Management 25, 357–384 (1999)
8. Zanko, M., Dawson, P.: Occupational Health and Safety Management in Organizations. A Review. International Journal of Management Reviews 14, 328–344 (2012)
9. Kumari, P., Mathew, L., Syal, P.: Increasing trend of wearables and multimodal interface for human activity monitoring: A review. Biosensors & bioelectronics 90, 298–307 (2017)
10. Della Mea, V.: What is e-Health (2): The death of telemedicine? Journal of Medical Internet Research 3, e22 (2001)
11. Lupton, D.: Self-tracking, health and medicine. Health Sociology Review 26, 1–5 (2016)
12. Dennison, L., Morrison, L., Conway, G., Yardley, L.: Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. Journal of Medical Internet Research 15, e86 (2013)
13. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., Strub, H.B.: Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. Proceedings of the 8th International Conference on Ubiquitous Computing, 261–278 (2006)
14. Abiresearch: mHealth Wearables Help Employers Achieve Higher Corporate Wellness Participation Rates | ABI Research, https://www.abiresearch.com/press/mhealth-wearables-help-employers-achieve-higher-co/
15. Cugelman, B.: Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 1 (2013)
16. O'brien, T., Troutman-Jordan, M., Hathaway, D., Armstrong, S., Moore, M.: Acceptability of wristband activity trackers among community dwelling older adults. Geriatric Nursing 36, S21-S25 (2015)
17. Pfeiffer, J., Entress-Fuersteneck, M.v., Urbach, N., Buchwald, A.: Quantify-Me: Consumer acceptance of wearable self-tracking devices. ECIS, Research Paper 99 (2016)
18. Ahlan, A.R., Ahmad, B.I.: User acceptance of health information technology (HIT) in developing countries: a conceptual model. Procedia Technology 16, 1287–1296 (2014)
19. Dünnebeil, S., Sunyaev, A., Blohm, I., Leimeister, J.M., Krcmar, H.: Determinants of physicians' technology acceptance for e-health in ambulatory care. International journal of medical informatics 81, 746–760 (2012)
20. Deng, X., Christodoulidou, N.: Understanding User Values of Wearable Computing. ICIS 2015 Proceedings (2015)
21. Baig, M.M., Gholamhosseini, H.: Smart health monitoring systems: an overview of design and modeling. Journal of Medical Systems 37, 9898 (2013)
22. Gao, W., Emaminejad, S., Nyein, H.Y.Y., Challa, S., Chen, K., Peck, A., Fahad, H.M., Ota, H., Shiraki, H., Kiriya, D.: Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 529, 509–514 (2016)
23. Pantelopoulos, A., Bourbakis, N.G.: A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Trans. Syst., Man, Cybern. C 40, 1–12 (2010)

24. Piwek, L., Ellis, D.A., Andrews, S., Joinson, A.: The Rise of Consumer Health Wearables: Promises and Barriers. PLoS medicine 13, e1001953 (2016)
25. Seshadri, D.R., Drummond, C., Craker, J., Rowbottom, J.R., Voos, J.E.: Wearable Devices for Sports. IEEE pulse 8, 38–43 (2017)
26. Curone, D., Secco, E.L., Tognetti, A., Loriga, G., Dudnik, G., Risatti, M., Whyte, R., Bonfiglio, A., Magenes, G.: Smart garments for emergency operators. IEEE Transactions on Information Technology in Biomedicine 14, 694–701 (2010)
27. Eysenbach, G.: What is e-health? Journal of Medical Internet Research 3, E20 (2001)
28. Lupton, D.: M-health and health promotion: The digital cyborg and surveillance society. Social Theory & Health 10, 229–244 (2012)
29. Gao, L., Bai, X.: A unified perspective on the factors influencing consumer acceptance of internet of things technology. Asia Pacific Journal of Marketing and Logistics 26, 211–231 (2014)
30. Lupton, D.: Health promotion in the digital era: a critical commentary. Health Promotion International 30, 174–183 (2015)
31. Weinstein, N.D.: Testing four competing theories of health-protective behavior. Health psychology: official journal of the Division of Health Psychology, American Psychological Association 12, 324–333 (1993)
32. Becker, M.H.: The health belief model and personal health behavior. Health education monographs 2, 324–473 (1974)
33. Glanz, K., Rimer, B.K., Viswanath, K.: Health behavior and health education: theory, research, and practice. John Wiley & Sons (2008)
34. Rogers, R.W.: A protection motivation theory of fear appeals and attitude change. The journal of psychology 91, 93–114 (1975)
35. Sun, Y., Wang, N., Guo, X., Peng, Z.: Understanding the acceptance of mobile health services: a comparison and integration of alternative models. Journal of Electronic Commerce Research 14, 183 (2013)
36. Jayanti, R.K., Burns, A.C.: The antecedents of preventive health care behavior: An empirical study. Journal of the Academy of Marketing Science 26, 6 (1998)
37. Ajzen, I.: The theory of planned behavior. Organizational behavior and human decision processes 50, 179–211 (1991)
38. Taylor, S., Todd, P.A.: Understanding Information Technology Usage. A Test of Competing Models. Information Systems Research 6, 144–176 (1995)
39. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Quarterly 27, 425–478 (2003)
40. Ajzen, I., Fishbein, M.: Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley (1975)
41. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319 (1989)
42. Amberg, M., Hirschmeier, M., Wehrmann, J.: The Compass Acceptance Model for the analysis and evaluation of mobile services. IJMC 2, 248 (2004)
43. Melzner, J., Heinze, J., Fritsch, T.: Mobile Health Applications in Workplace Health Promotion. Procedia Technology 16, 1374–1382 (2014)
44. Fagan, M.H., Neill, S., Wooldridge, B.R.: Exploring the intention to use computers: An empirical investigation of the role of intrinsic motivation, extrinsic motivation, and perceived ease of use. Journal of Computer Information Systems 48, 31–37 (2008)
45. Igbaria, M., Iivari, J., Maragahh, H.: Why do individuals use computer technology? Information & Management 29, 227–238 (1995)

46. Karaiskos, D., Kourouthanassis, P., Giaglis, G.: User Acceptance of Pervasive Information Systems. ECIS 2007 Proceedings (2007)
47. Yang, H., Yu, J., Zo, H., Choi, M.: User acceptance of wearable devices: An extended perspective of perceived value. Telematics and Informatics 33, 256–269 (2016)
48. Wirtz, B.W., Mory, L., Ullrich, S.: eHealth in the public sector. Public administration 90, 642–663 (2012)
49. Cocosila, M., Archer, N.: Adoption of mobile ICT for health promotion: an empirical investigation. Electronic Markets 20, 241–250 (2010)
50. Hung, M.-C., Jen, W.-Y.: The adoption of mobile health management services: an empirical study. Journal of Medical Systems 36, 1381–1388 (2012)
51. Bhattacherjee, A., Hikmet, N.: Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. European Journal of Information Systems 16, 725–737 (2007)
52. Arning, K., Ziefle, M., Arning, J.: Comparing apples and oranges? Exploring users' acceptance of ICT-and eHealth-applications. Digital camera 83, 0–1 (2008)
53. Ryan, Deci: Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary educational psychology 25, 54–67 (2000)
54. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 36, 157–178 (2012)
55. Brown, S.A., Venkatesh, V.: Model of adoption of technology in households. MIS Quarterly, 399–426 (2005)
56. Kuvaas, B.: Work performance, affective commitment, and work motivation. The roles of pay administration and pay level. Journal Organizational Behavior 27, 365–385 (2006)
57. Bruner, G.C., Kumar, A.: Gadget lovers. Journal of the Academy of Marketing Science 35, 329–339 (2007)
58. Vallerand, R.J.: Toward a hierarchical model of intrinsic and extrinsic motivation. Advances in Experimental Social Psychology 29, 271–360 (1997)
59. Janz, N.K., Becker, M.H.: The Health Belief Model: a decade later. Health Education Quarterly 11, 1–47 (1984)
60. Guo, X., Zhang, X., Sun, Y.: The privacy–personalization paradox in mHealth services acceptance of different age groups. Electronic commerce research and applications 16, 55–65 (2016)
61. Or, C.K.L., Karsh, B.-T.: A Systematic Review of Patient Acceptance of Consumer Health Information Technology. Journal of the American Medical Informatics Association : JAMIA 16, 550–560 (2009)
62. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Information Systems Research 2, 192–222 (1991)
63. Joshi, A., Kale, S., Chandel, S., Pal, D.: Likert Scale. Explored and Explained. British Journal of Applied Science & Technology 7, 396–403 (2015)
64. Kankanhalli, A., Tan, B.C.Y., Wei, K.-K.: Contributing knowledge to electronic knowledge repositories: an empirical investigation. MIS Quarterly, 113–143 (2005)
65. Moon, J.-W., Kim, Y.-G.: Extending the TAM for a World-Wide-Web context. Information & Management 38, 217–230 (2001)
66. Guay, F., Vallerand, R.J., Blanchard, C.: On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and Emotion 24, 175–213 (2000)

67. Bock, G.-W., Zmud, R.W., Kim, Y.-G., Lee, J.-N.: Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. MIS Quarterly, 87–111 (2005)
68. Lau, R.R., Ware Jr, J.F.: Refinements in the Measurement of health-specific locus-of-control beliefs. Medical care 19, 1147–1158 (1981)
69. Moorman, C., Matulich, E.: A Model of Consumers' Preventive Health Behaviors: The Role of Health Motivation and Health Ability. J Consum Res 20, 208–228 (1993)
70. Champion, V.L.: Instrument development for health belief model constructs. Advances in Nursing Science 6, 73–85 (1984)
71. Champion, V.L.: Revised susceptibility, benefits, and barriers scale for mammography screening. Research in Nursing & Health 22, 341–348 (1999)
72. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC). The Construct, the Scale, and a Causal Model. Information Systems Research 15, 336–355 (2004)
73. Urbach, N., Ahlemann, F.: Structural Equation Modeling in Information Systems Research Using Partial Least Squares. Journal of Information Technology Theory and Application (JITTA) 11 (2010)
74. Henseler, J.: PLS-MGA: A non-parametric approach to partial least squares-based multi-group analysis. In: Challenges at the interface of data analysis, computer science, and optimization, pp. 495–501. Springer (2012)

Understanding Patient Pathways in the Context of Integrated Health Care Services - Implications from a Scoping Review

Peggy Richter[1], Hannes Schlieter[1]

[1] Technische Universität Dresden, Chair of Wirtschaftsinformatik, esp. Systems Development, Dresden, Germany
{peggy.richter2,hannes.schlieter}@tu-dresden.de

Abstract. Healthcare systems in western countries are continuously working to achieve efficient resource allocation and to improve access to quality medical care. The implementation of standardised care processes promises better integration and coordination of care across several healthcare providers. In this context, an increasing use of the term patient pathway is recognised within official documents provided by health authorities and within scientific publications in recent years. However, a common understanding, distinguishing the term from other pathway approaches such as care- or clinical pathways, is missing. By means of a scoping review we analysed 132 publications in order to clarify key concepts and the understanding of patient pathways. Six common themes in the literature were identified and results show that individualisation and care continuity are essential descriptive characteristics. Using this motivation, we discuss the main implications for research and practice by the example of comprehensive cancer care in the European Union.

Keywords: patient pathways, care networks, scoping review

1 Introduction

Health care faces a broad spectrum of transition processes that necessitate integrated care delivery. In this context, demographic change, skilled worker shortage and an increasing number of patients with multimorbidity and chronic diseases are among the main drivers [1, 2]. For the latter, cancer is one of the most common and costly diseases in western countries [3, 4]. In order to coordinate cancer care on the national level and to increase access to quality cancer care, the implementation of Comprehensive Cancer Care Networks (CCCNs) is recommended by the European guide on quality improvement in comprehensive cancer control [5]. Such networks integrate different institutions and institutional units representing all relevant episodes for a patient's cancer care, i. e. research, prevention, diagnosis, treatment, follow-up, rehabilitation and end-of-life care [5]. One of the CCCNs' tasks is the provision of practical support tools. In this context, comprehensive, integrated patient pathways are recognised as a valuable approach [5]. Whereas the term patient pathway is often used with regard to

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

optimising cancer care processes and aligning information and communication flows, a common terminological basis is still missing. This has negative impact on the harmonisation of such big scale activities, in this case on EU level, and on the communication of their maturity in general. Hence, to further advance the utilisation of patient pathways in cancer care and beyond, clarification of the concept is necessary. It is still unclear whether patient pathways are any different from already well-established pathway approaches such as care pathways or clinical pathways. According to the definition used by the European Pathway Association, a "care pathway is a complex intervention for the mutual decision making and organisation of care processes for a well-defined group of patients during a well-defined period" [6]. Clinical pathways particularly focus on the care provision within a single institution, e. g. a hospital [7].

The article aims to examine the literature body available on patient pathways. On this basis, key concepts of patient pathways shall be clarified and implications for future research and practice shall be discussed. Therefore, three research questions are to be answered: (RQ1) How has the literature on patient pathways developed over the years and which themes are addressed in the literature? (RQ2) What are characteristics of patient pathways including characteristics that differentiate them from other pathway approaches? (RQ3) What are potential implications for practice and future research? Accordingly, the article is structured as follows: The scoping review method used to address the research questions is described in section 2. The results are presented in section 3 by describing identified themes within the patient pathway literature. Also, common characteristics of patient pathways are derived within this section. The results of the review are discussed in the context of current literature and practice in section 4. This also includes the discussion of implications for research and practice as well as limitations of our study. A conclusion is given in section 5 by summarising the results and the contributions of the presented work.

2 Method

2.1 Scoping Review

In order to answer the research questions, a scoping review was conducted. Unlike a systematic literature review, a scoping review is a review type which is used to map key concepts underpinning a research area [8, 9]. A common purpose of scoping reviews is the identification of topics for future research [10]. They are often performed to determine and represent the body of literature and available evidence on a topic [8]. Since there is yet no comprehensive literature review about patient pathways available, the conduction of a scoping review is the appropriate choice to answer our research questions. Also, the research objectives are among those, that Anderson et al. (2008) state as the key criteria for which a scoping review is reasonable, e. g. clarification of the conceptual understanding of a topic where definitions are unclear and identification of research gaps advising on future research [10].

The conduction and reporting of the scoping review follows the guidelines proposed by Peters et al. [9]. The process of developing a review protocol as well as the literature search and selection process are not different from a systematic literature review, except that there is no formal assessment of the methodological quality of the included literature [8, 9]. The developed review protocol included the definition of the search strategy, search terms, databases, screening approach, and the inclusion criteria.

In order to clarify the characteristics of the review, we draw on the established taxonomy for literature reviews described by Cooper [11] and define the focus, goal, perspective, coverage, organisation and audience as intended with the scoping review on patient pathways. The focus areas are outcomes, theories and applications. With respect to the research objective of this article, the goal of the review is twofold, i. e. it aims at the identification of central issues (see RQ1 and RQ2) as well as at the integration of existing literature (see RQ3). We take a neutral perspective in the presentation of the results and cover literature exhaustively with selected citations due to space limitations. The review is organised conceptually. The intended audiences are both scholars and practitioners in the fields of information systems and health care.

2.2 Search Strategy

The search strategy comprised a scientific database search and a google search in order to include both scientific articles and grey literature (e. g. government, business or institution reports regarding patient pathways) [12, 13]. The searches were carried out in April and May 2018. The search and review process is depicted in Figure 1.

During the screening phase, the exclusion criteria were non-scientific publication types (e. g. letters to the editors), the unavailability of an English abstract (reasonable for the results of the database search), or an extraneous topic (e. g. disease specific research aims that did not have patient pathways as a central topic). The high number of unavailable database records mostly results from unavailable poster abstracts. In such cases, the authors were contacted and asked to share their publication.

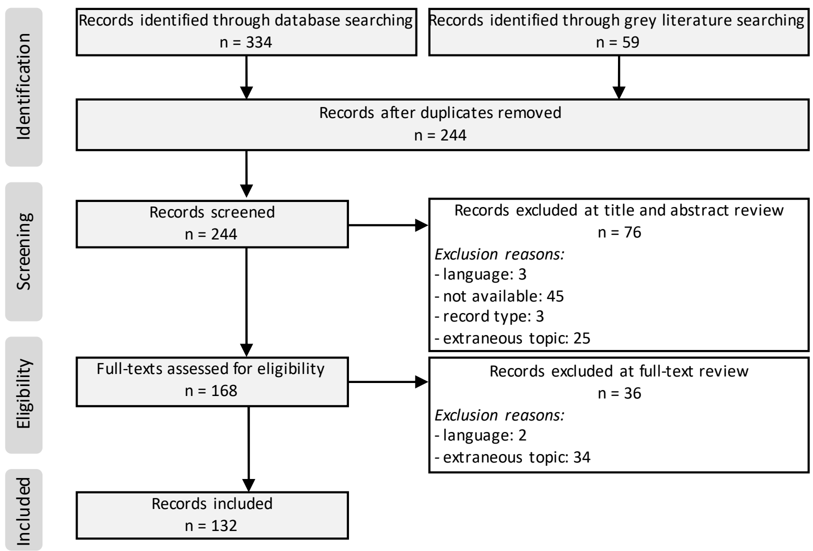


Figure 1. Scoping review process (representation mode based on [14])

During the full-text assessment, we did not restrict the analysis to full papers but also included abstracts, poster abstracts and posters, since one aim of the scoping review is the analysis of the currently available literature body on patient pathways. Included full-texts were in English or German, the latter occurring twice.

2.3 Database Search

We searched for "patient pathway" OR "patient pathways" in the titles of articles in PubMed, EbscoHost Academic Search Complete, Web of Science, and ScienceDirect. The decision to exclusively search for the term patient pathway is based on initial search term tests. These tests included potential synonyms such as clinical-, treatment-, care-, or integrated pathways and tested combinations with terms related to “definition” or “development”. The searches resulted in a large amount of irrelevant articles and since the focus of the scoping review is the examination of the term patient pathway, we decided to exclusively search for this term.

The results were screened based on titles and abstracts by four reviewers. This procedure meets the requirement for at least two reviewers necessary for a scoping review [9]. In order to create a common understanding for the screening process a pre-test was conducted. Each reviewer analysed the first 15 included records regarding the context in which the term patient pathway was used. In a consensus meeting, the findings were discussed and the authors decided on a preliminary classification of research themes in the literature on patient pathways. This classification was assessed and refined in a second pre-test by applying it to the first 25 records of the included records and another consensus meeting. Based on this revised classification, the authors mapped all included records. During this, minor iterative refinements of the classification took place, leading to its final version as described in section 3.1.

2.4 Search for Grey Literature

We searched google exclusively for the singular and plural of the term patient pathway, for the same reasons as described for the database search. The first six results pages (the first 59 results, excluding hits to language translation pages) were screened by the authors. After that, there were scarcely any more relevant hits, which justifies ending the search at this point. The literature classification developed during the pre-test with the database search results was also applied to the included grey literature.

3 Results

3.1 Themes in the Patient Pathway Literature

As a first, theme-independent result, it can be stated that the interest in the examination of patient pathways in research and practice has increased over the years. This is shown by the increase in publications per year as depicted in Figure 2.

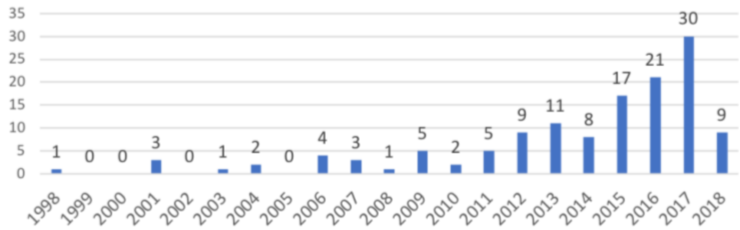


Figure 2. Number of publications per year (n = 132)

The literature can be roughly divided into publications examining patient pathways in general (n = 39) and publications addressing patient pathways for certain diseases (n = 93). For the latter, a major focus is on patient pathways in cancer care. As shown in Figure 3 (a), patient pathways for tuberculosis, heart diseases or alcoholism/drug addiction were examined much less frequently. The rest of the disease-specific literature was highly diverse. Thus, all diseases that were addressed only once or twice in the analysed patient pathway literature were summarised as others.

As described in section 2.3, during the analysis of the titles, abstracts and full-texts of the searched literature, a classification of six common research themes was developed and iteratively refined during the process. The identified themes regarding patient pathways are: definition and conceptualisation, development and implementation, analysis of patient pathways, responsibilities and roles, tool- and IT-support, and simulation. The frequency distribution of the themes is depicted in Figure 3 (b). Each theme is described in detail in the following.

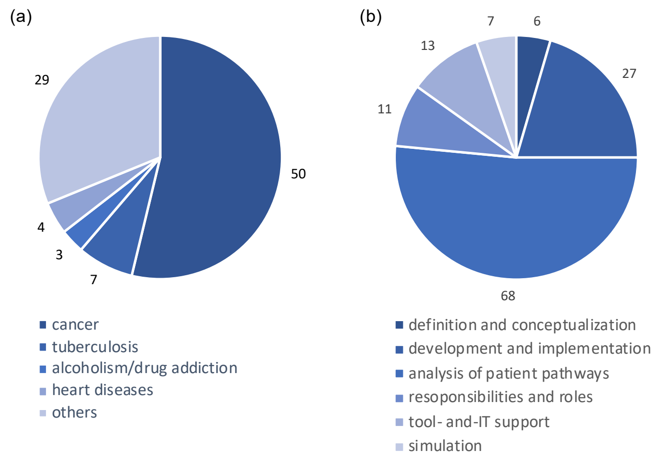


Figure 3. Results of the analyses: (a) focus of disease-specific patient pathway literature (n = 93), (b) themes in the patient pathway literature (n = 132)

Definition and Conceptualisation. Currently, there are only few articles (n = 6) examining characteristics or the notion of patient pathways. However, there is neither a comprehensive analysis of the patient pathway approach nor a common definition available. Instead, the current literature discusses single important issues in relation to the patient pathway approach. For example, Berntsen and colleagues [15] qualitatively analysed the goals of care in individual patient pathways and concluded that the alignment of functional, biological and personal goals in a network of multidisciplinary care providers is essential to support care continuity. Another issue, addressed by Salamonsen et al. [16], is the influence of disruptive health and life events on the patient pathways. Also, they argue that patient pathways are beyond the known concept of clinical pathways, because they are not solely understood as the standardised, guideline-based provision of health care for certain patient types but also address multimorbidity and patient preferences [16]. Further basic characteristics of patient pathways can be taken from the patient pathway diagram as part of the NHS Data Model and Dictionary [17]. It describes the attributes and relationships to health care activities and organisations for the patient pathway class.

Development and Implementation. There are 27 publications that address the issues of development and implementation of patient pathways. We also included publications that present concrete patient pathways for certain diseases and health situations in this theme because they indicate how patient pathways are represented in practice. The concrete patient pathways analysed for the scoping review were represented as checklists [18], flowcharts [19] or non-standardised semi-formal process models often combined with textual descriptions [20, 21]. Standards, clinical practice guidelines or the results of systematic literature reviews are the starting points for pathway development and can be complemented by expert knowledge or experiences [22, 23]. Wicke et al. [24] propose a construction process for patient pathways based on a lifecycle consisting of four phases, i. e. preparation, construction and testing (main phase), implementation, and maintenance (controlling and revision). In the context of the latter, the collection of quality, service improvement and redesign tools could be used to improve certain stages of patient pathways [25]. So far, patient pathway development is rather addressed in the context of single health care institutions (e. g. a hospital [24]) than on network level.

The development and implementation are tasks of a multidisciplinary, inter-organisational team consisting of all key stakeholders along the patient pathway, e. g. involving hospital and community staff [23, 24, 26]. The development process should include consensus team meetings in order to reach agreement between all partners [27, 28]. A prominent example for nationally implemented patient pathways are the Danish Cancer Patient Pathways that were developed with a consensus seeking model ensuring the involvement and cooperation between bureaucrats, health professionals and politicians [29]. Besides the involvement of professional and administrative staff in the development process, the patient's perspective is highlighted as well. Focus group discussions are one way to identify patients' preferences and to integrate their input in the pathway development and maintenance phases [30].

Analysis of Patient Pathways. The analysis of patient pathways was subject to more than half of the publications analysed (n = 68). This theme can be further differentiated into four sub-topics, which are (I) patient pathway analysis (n = 16), (II) the analysis of patient pathway usage (n = 27), (III) the analysis of patient pathway effects (n = 23), and (IV) the examination of data sources to be used for analysis purposes (n = 2). Patient pathway analysis (I) is an established phrase being used to describe the alignment between the care seeking patterns of patients with a certain disease and the availability of corresponding health care services [31, 32]. In this understanding, the patient pathway is not a predefined, standardised process but the actual, unplanned journey of a patient seeking health care services to address her/his health conditions. Based on this understanding, articles examining the pathways of patients before and after hospitalisation [33] or how patients decide where to seek care and what factors influence this decision [34] were also included in this sub-topic. In contrast, the analysis of the practical usage of patient pathways (II) is based upon predefined, standardised pathways. This sub-topic includes the analysis of patient characteristics for a certain patient pathway [35, 36], pathway compliance analyses [37], and process analyses (e. g. sources of delay [38, 39] or utilisation of certain activities in a pathway [40]). The analysis of patient pathway effects (III) comprises outcome studies, i.e. the analysis of the results after patient pathway implementation. Here, patient-related outcomes, such as survival and mortality, complications, quality of life and treatment-related side-effects [41, 42], and organisation-related outcomes, such as waiting times, length of hospital stay or timeliness of care [43, 44], are the focus. Another, rather minor sub-topic is the examination of data sources that can be used for analysis purposes (IV). In this context, the need for data linkage of electronic health records with clinical data registries as well as across regional borders is highlighted [45, 46]. This would increase the understanding of the complete patient pathway covering multiple involved care providers and also support national research.

Responsibilities and Roles. The responsibilities and roles of different stakeholder groups are discussed in 11 of the analysed articles. The role of nurses turned out be particularly important for patient navigation and early patient pathway optimisation, i.e. coordinating interventions, streamlining and planning the pathway, tailoring interventions to individual patient's needs and preferences [47–49]. The literature examines the relevance of nurses for patient management and for the provision of seamless care in both the acute care (clinical nurse specialists [47, 48]) and non-acute care setting (community nurses [49]). Other roles and their impact on patient pathways addressed in the literature are emergency care practitioners [50] and community pharmacists [51]. The import role of patients themselves is reflected with the patient-centred approach called user-led health care. It describes the systematic involvement of patients in the planning and execution of their individual treatment and care process and aligns patients' responsibilities with their preferences [52].

Tool- and IT-Support. Improving the utilisation and the streamlining of patient pathways by tools and information technology is a central issue for 13 of the analysed literature records. Here, patient education and patient empowerment by the online provision of information materials are typical applications. For example, there are internet-based, interactive patient pathways used as education tools for breast cancer patients [53] or to explain a forthcoming surgical journey [54], and interactive, individualised online patient pathways with detailed, personalised timelines to keep patients informed and involved [55]. Another topic in relation to tool- and IT-support is the use of telemedicine in order to streamline a patient pathway, e. g. with early tele-assessment of stroke patients [56]. Furthermore, there are triage tools to assess and allocate patients to the appropriate pathway [57] and technical solutions for tracking patient pathways [58].

Simulation. The literature addressing simulation of patient pathways (n = 7) applies it for example to predict their outcomes. Simulation models are used to produce quantified output of patients with a certain diagnosis [59] or to simulate scenarios at the population level in order to support intervention planning of public health care [60]. Furthermore, simulation is used for teaching and training purposes, e. g. by fully simulating a hospital across the entire patient pathway in order to train expert health care providers [61] or by simulating the surgical patient pathway for undergraduate students to supplement classroom medicine and clinical practice [62].

3.2 Characteristics of patient pathways

The scoping review results show that currently a common definition of patient pathways is not available. However, characteristics of this approach can be identified from the included literature by analysing their understanding of patient pathways. It was found that there are related pathway terms also used as synonyms for patient pathways. Mainly, these are care pathways [47, 49, 63–65], treatment pathways [47, 66, 67], and patient journeys [45, 68, 69]. However, the usage of these terms was not reasoned in the articles. The full-texts of the included literature were screened for statements describing characteristics of patient pathways, such as:

"[…] 'patient pathways' from a patient perspective, understood as incorporated into socioculturally constructed life courses. […] Not only 'health events', but also 'life events' are included in our understanding of patient pathways." [16]

"Based on treatment guidelines, patient pathways display an optimal sequence of staff actions in the preoperative, operative, and postoperative in- and outpatient treatment." [24]

"Patient pathways are tools that assist in providing general guidelines of care for dealing with individuals and groups of patients suffering from a wide variety of diseases." [70]

Based on the analysis of such statements and the identified themes in the literature, descriptive characteristics of patient pathways can be summarised as follows. Patient pathways are:

- stating and aligning functional, biological, and patient-related goals of care [15, 16],
- focusing on patient group and individual patient planning and -management for complex long-term conditions [16, 70–72],
- describing and sequencing key components of care to guide care provision and the patient journey [18–21, 23, 24],

- comprising the whole route a patient takes including inpatient and outpatient settings and thus, are typically inter-organisational pathways [45, 68, 71, 73] but individual stakeholders can focus on single episodes (e. g. surgery or hospital stay) [24]
- developed, implemented and used by a multidisciplinary care team consisting of professional and informal caregivers and involving the patient [23, 24, 26, 30, 52],
- evidence-based (medical guidelines, standards) and experts' experiences [22, 23],
- used for patient information, documentation, monitoring and evaluation purposes (e. g. assessment of quality and efficiency of care delivery or patient-related outcome measures) [37–44, 53, 55, 74].

4 Discussion

As the analysis revealed, more than half of the reviewed publications are related to oncological diseases. This emphasises the initial statement that patient pathways are recognised as a valuable approach in cancer care in order to create seamless care, inform the patient, plan the care process, and implement medical guidelines [5]. The results of the scoping review also show that the majority of publications addresses the analysis of various aspects related to patient pathways such as the effects of their application or their usage (see Figure 3 (b)). Still, the broad use of the term indicates some ambiguity in its understanding. Surprisingly, there are a lot of papers analysing effects of patient pathways but without referring to a common definition of this concept.

Although some papers discuss the use of patient pathway in an intra-organisational context and other focus on the inter-organisational setting, the proposed characteristics cover both aspects because patient pathways within a single institution are also embedded in the inter-organisational route of a patient. Interfaces and the relation to the overall process have to be described accordingly. Furthermore, ambiguity exists regarding the time perspective of patient pathways. Some authors focus more on the general journey of patients through the health care delivery system, analysed retrospectively for example by means of patient pathway analysis. Others rather focus on prospectively defined pathways for a patient (group) to guide the provision of care. The latter highlights the planning character of a pathway, which is also the aim of related pathway approaches such as clinical pathways and care pathways.

In the context of existing literature on other pathway approaches, patient pathways comprise the core concept of care pathways, that are already well established in the field of medical process management nowadays [75, 76]. However, patient pathways have a stronger focus on the individual patient. This is particularly apparent in form of aligning the goals of care to a patient's needs and preferences and, accordingly, tailoring the pathway to the individual. The concept of patient pathways emphasises the care process from the perspective of the patient and also includes mechanisms of empowerment and engagement. Based on the identified characteristics, patient pathways are also not equivalent to clinical pathways [7] because these typically do not cover the whole care chain across in- and outpatient settings.

The current state of literature provides several action points for future research. A consensus regarding the patient pathway understanding should be developed, e. g. by an expert panel discussing and rating the identified characteristics. From a practical perspective, a stronger methodological support for developing and implementing patient pathways is desirable. This could be achieved by providing a tool for the preparation of patient pathways, their adaption to patient-individual characteristics and to local specifics of the application environment. Latest research, e. g. on adaptive and personalised pathways [77] or on multi-perspective pathway modelling languages [78], can be exploited for this purpose. The mechanisms of reference modelling [79] might provide appropriate means to combine generic templates with intended governance for health process design. Taking up the example of comprehensive cancer care, a method for patient pathway development could allow CCCNs to prepare common templates of patient pathways for specific tumour entities. These could for example specify the main goals, phases and milestones of care for a specific patient type as well as roles and tasks within the care network. Such a template, functioning as a reference model, could then be adapted to regional conditions and patient individualities based on adaption guidelines provided along with the method. In order to increase usability, the specification of user requirements is essential, since the users have an application-oriented background being e. g. physicians, nurses or a cancer patient. In perspective, this approach could increase comparability and set process-oriented quality standards for cancer care in CCCNs. This could contribute to developing more process-oriented measures for quality and performance assessment. Thereby, insights for continuous patient pathway improvement could be gained.

Critically reflecting on the methodological approach of the presented paper, the scoping review could be broadened by expanding the database search to titles and abstracts and by adding forward and backward searches. The mapping of the literature could be enriched by a multi-disciplinary review team [10] that also involves scholars from the health care domain. Due to the nature of the scoping review it does not include a process of quality assessment of the literature and thus, has some limitations [8]. For this reason, recommendations for practice are preliminary and the understanding of patient pathways could be further examined by a systematic literature review. The conducted scoping review determined the high value and scope of such.

5 Conclusion

The article aimed at examining the current literature body available on patient pathways in order to clarify key characteristics of this approach and to discuss implications for research and practice. A systematic scoping review was conducted. It included a database and a grey literature search. An increase in the discussion of the concept in the literature was confirmed. There were six common themes identified in the current literature on patient pathways. These are definition and conceptualisation, development and implementation, analysis of patient pathways, responsibilities and roles, tool- and IT-support, and simulation. The majority of publications addresses analytical topics, such as the analysis of patient pathway usage or effects but a common definition is currently not available. However, there are typical characteristics of patient pathways that were summarised based on the reviewed literature. Patient pathways differ from other pathway approaches as they rather aim at planning care for multimorbid patients with complex health conditions. They focus on the inter-organisational setting and patients' needs and preferences. We also discussed potential implications for practice and research. Particularly, the methodological and technological support for developing and implementing patient pathways should be improved.

The presented work contributes to the knowledge base by consolidating the understanding of patient pathways and by summarising typical characteristics. As we pointed out, the patient pathway approach builds upon the core concept of care pathways. Since there are concerns regarding the feasibility of care pathways for complex health conditions that require integrated care, we see the opportunity that patient pathways drive a transition to a broader utilisation of pathways for chronic and complex health conditions such as cancer. In summary, our work informs the debate on patient pathways and is a starting point for future research and practical applications.

6 Acknowledgements

This article was funded by the European Union's Health Programme (2014-2020). We thank Carola Gißke, Roy Stöhr and Stephanie Krug for participating in the screening of literature and in the pre-test.

Co-funded by the Health Programme of the European Union

iPAAC INNOVATIVE PARTNERSHIP FOR ACTION AGAINST CANCER

References

1. Hujala, A., Taskinen, H., Rissanen, S.: Policy Brief 26. How to support integration to promote care for people with multimorbidity in Europe? (2016)
2. Minkman, M.: The current state of integrated care: an overview. J. Integr. Care. 20, pp. 346–358 (2012)
3. Busse, R., Blümel, M., Scheller-Kreinsen, D., Zentner, A.: Tackling chronic disease in Europe. Strategies, interventions and challenges. Kopenhagen (2010)
4. Banks, J., Muriel, A., Smith, J.P.: Disease prevalence, disease incidence, and mortality in the United States and in England. Demography. 47, pp. S211–S231 (2010)
5. Albreht, T., Kiasuwa, R., Van der Bulcke, M. eds: EUROPEAN guide on quality improvement in comprehensive cancer control. Ljubljana: National Institute of Public Health; Brussels: Scientific Institute of Public Health (2017)
6. European Pathway Association: Care Pathways, http://e-p-a.org/care-pathways/ (Accessed: 23.05.2018)
7. Kinsman, L., Rotter, T., James, E., Snow, P., Willis, J.: What is a clinical pathway? Development of a definition to inform the debate. BMC Med. 8, pp. 31–33 (2010)
8. Grant, M.J., Booth, A.: A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf. Libr. J. 26, pp. 91–108 (2009)
9. Peters, M.D.J., Godfrey, C.M., Khalil, H., Mcinerney, P., Parker, D., Soares, C.B.: Guidance for conducting systematic scoping reviews. Int. J. Evid. Based Healthc. 13, pp. 141–146 (2015)

10. Anderson, S., Allen, P., Peckham, S., Goodwin, N.: Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. Health Res. Policy Syst. 6, (2008).
11. Cooper, H.M.: Organizing knowledge syntheses: A taxonomy of literature reviews. Knowl. Soc. 1, 104–126 (1988).
12. Aromataris, E., Riitano, D.: Constructing a search strategy and searching for evidence. A guide to the literature search for a systematic review. Am. J. Nurs. 114, 49–56 (2014).
13. Tillett, S., Newbold, E.: Grey literature at The British Library: revealing a hidden resource. Interlend. Doc. Supply. 34(2), pp. 70–73 (2006).
14. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 339, (2009).
15. Berntsen, G.K.R., Gammon, D.B., Steinsbekk, A., Foss, N., Ruland, C., Fonnebo, V.: What are the goals of care in individual Patient Pathways? An qualitative analysis of how different health concepts may affect goals for care. Int. J. Integr. Care. 15, e192 (2015)
16. Salamonsen, A., Kiil, M.A., Kristoffersen, A.E., Stub, T., Berntsen, G.R.: "My cancer is not my deepest concern": life course disruption influencing patient pathways and health care needs among persons living with colorectal cancer. Patient Prefer. Adherence. 10, pp. 1591–1600 (2016)
17. National Health Service (NHS Digital): Patient Pathway Diagram, https://www.datadictionary.nhs.uk/data_dictionary/diagrams/diagrams/patient_pathway_diagram_fr.asp?shownav=1 (Accessed: 24.05.2018)
18. Regina Qu'Appelle Health Region: Laryngectomy Patient Pathway, http://www.rqhealth.ca/rqhr-central/files/0537.pdf (Accessed: 24.05.2018)
19. Kingston, M., Smurthwaite, D., Dixon, S.: A pragmatic patient pathway ensuring appropriate safeguarding decisions for children with genital warts. Sex. Transm. Infect. 91, p. A96 (2015)
20. Cambridgeshire and Peterborough Clinical Commissioning group: Chronic back pain, https://www.cambridgeshireandpeterboroughccg.nhs.uk/health-professionals/patient-pathways/pain-management/chronic-back-pain/ (Accessed: 24.05.2018)
21. Cancer Care Ontario: Lung Cancer Diagnosis Patient Pathway, http://www.tbrhsc.net/wp-content/uploads/2017/03/CCO-Lung-Patient-Pathways_2013.pdf (Accessed: 24.05.2018)
22. Akdag, H.C., Canturk, N.Z.: Improvement of Breast Cancer Patient Pathway Using EUSOMA Standards and European Guidelines. Chir. Buchar. Rom. 1990. 112, pp. 449–456 (2017)
23. Walsh, D.A., Kelly, C., Bosworth, A., Price, C., Burbage, G.: Provisional guidelines for applying the Department of Health (England) 18-week-patient pathway to specialist rheumatology care. Rheumatol. Oxf. Engl. 46, pp. 1200–1206 (2007)
24. Wicke, C., Teichmann, R., Holler, T., Rehder, F., Becker, H.D.: Design and use of patient pathways in general surgery. Chir. Z. Alle Geb. Oper. Medizen. 75, pp. 907–915 (2004)
25. ACT Academy NHS Improvement: Quality, service improvement and redesign (QSIR) tools by stage of the patient pathway, https://improvement.nhs.uk/resources/quality-service-improvement-and-redesign-qsir-tools-stage-patient-pathway/ (Accessed: 24.05.2018)
26. Deery, P., Faithfull, S.: Developing a patient pathway to deliver a new oral chemotherapy. Prof. Nurse Lond. Engl. 19, pp. 102–106 (2003)
27. Hargreaves, J.A.: Methods for Determining Patient Pathway and Resource Use In Complex Abdominal Wall Reconstruction. Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res. 18, p. A706 (2015)
28. Greenwood, J.E.: Development of patient pathways for the surgical management of burn injury. ANZ J. Surg. 76, pp. 805–811 (2006)

29. Probst, H.B., Hussain, Z.B., Andersen, O.: Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians--a national Danish project. Health Policy Amst. Neth. 105, pp. 65–70 (2012)
30. Nemeth, L., Hendricks, H., Salaway, T., Garcia, C.: Integrating the patient's perspective: patient pathway development across the enterprise. Top. Health Inf. Manage. 19, pp. 79–87 (1998)
31. Garfin, C., Mantala, M., Yadav, R., Hanson, C.L., Osberg, M., Hymoff, A., Makayova, J.: Using Patient Pathway Analysis to Design Patient-centered Referral Networks for Diagnosis and Treatment of Tuberculosis: The Case of the Philippines. J. Infect. Dis. 216, pp. S740–S747 (2017).
32. Hanson, C., Osberg, M., Brown, J., Durham, G., Chin, D.P.: Finding the Missing Patients With Tuberculosis: Lessons Learned From Patient-Pathway Analyses in 5 Countries. J. Infect. Dis. 216, pp. S686–S695 (2017)
33. Sutiono, A.B., Faried, A., McAllister, S., Ganefianty, A., Sarjono, K., Arifin, M.Z., Derrett, S.: The Bandung neurosurgery patient outcomes project, Indonesia (Part II): Patient pathways and feasibility and acceptability of telephone follow-up. Int. J. Health Plann. Manage, pp. e49-e56 (2017)
34. Gage-Bouchard, E.A., Rodriguez, E.M., Saad-Harfouche, F., Miller, A., Erwin, D.: Factors influencing patient pathways for receipt of cancer care at an. PloS One. 9, p. e110649 (2014)
35. Moseholm, E., Lindhardt, B.O.: Patient characteristics and cancer prevalence in the Danish cancer patient pathway for patients with serious non-specific symptoms and signs of cancer-A nationwide, population-based cohort study. Cancer Epidemiol. 50, pp. 166–172 (2017)
36. Lubman, D.I., Garfield, J.B.B., Manning, V., Berends, L., Best, D., Mugavin, J.M., Lam, T., Buykx, P., Larner, A., Lloyd, B., Room, R., Allsop, S.: Characteristics of individuals presenting to treatment for primary alcohol problems versus other drug problems in the Australian patient pathways study. BMC Psychiatry. 16, p. 250 (2016)
37. Lismont, J., Janssens, A.-S., Odnoletkova, I., Vanden Broucke, S., Caron, F., Vanthienen, J.: A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways. Comput. Biol. Med. 77, pp. 125–134 (2016)
38. Normahani, P., Standfield, N.J., Jaffer, U.: Sources of Delay in the Acute Limb Ischemia Patient Pathway. Ann. Vasc. Surg. 38, pp. 279–285 (2017)
39. Dyrop, H.B., Safwat, A., Vedsted, P., Maretty-Nielsen, K., Hansen, B.H., Jorgensen, P.H., Baad-Hansen, T., Bunger, C., Keller, J.: Cancer Patient Pathways shortens waiting times and accelerates the diagnostic process of suspected sarcoma patients in Denmark. Health Policy Amst. Neth. 113, pp. 110–117 (2013)
40. McCaughey, E.J., Li, L., Georgiou, A., Golding, M.H., Westbrook, J.I.: Imaging for patients presenting to an emergency department with back pain: Impact on patient pathway. Emerg. Med. Australas. EMA. 28, pp. 412–418 (2016)
41. Jensen, H., Torring, M.L., Vedsted, P.: Prognostic consequences of implementing cancer patient pathways in Denmark: a comparative cohort study of symptomatic cancer patients in primary care. BMC Cancer. 17, p. 627 (2017)
42. Ryhanen, A., Rankinen, S., Siekkinen, M., Korvenranta, H., Leino-Kilpi, H.: The impact of Breast Cancer Patient Pathway (BCPP) to patient's anxiety. Eur. J. Cancer. 49, p. S387 (2013)
43. Cook, C., Towler, L.: Quicker cancer care: reshaping patient pathways. (Cover story). Nurs. Manag. - UK. 16, pp. 20–23 (2009)
44. Laut, K.G., Foldspang, A.: The effects on length of stay of introducing a fast track patient pathway for myocardial infarction: a before and after evaluation. Health Serv. Manage. Res. 25, pp. 31–34 (2012)

45. Thompson, C.A., Kurian, A.W., Luft, H.S.: Linking electronic health records to better understand breast cancer patient pathways within and between two health systems. EGEMS Wash. DC. 3, p. 1127 (2015)
46. Boyd, J.H., Randall, S.M., Ferrante, A.M., Bauer, J.K., McInneny, K., Brown, A.P., Spilsbury, K., Gillies, M., Semmens, J.B.: Accuracy and completeness of patient pathways - the benefits of national data linkage in Australia. BMC Health Serv. Res. 15, pp. 312–319 (2015)
47. Basketter, V., Benney, M., Causer, L., Fleure, L., Hames, D., Jones, S., Patel, K., White, L.: Continuous, comprehensive and crucial care: The role of the CNS in the metastatic castration-resistant prostate cancer patient pathway. Br. J. Nurs. Mark Allen Publ. 27, pp. S1–S8 (2018)
48. Dunbar, J.L., Maddock, N., Hughes, J., McIver, A.: Optimising the patient pathway with early CNS intervention. Lung Cancer. 103, p. S42 (2017)
49. Bowers, B.: The role of community nurses in providing a seamless patient pathway. Br. J. Nurs. Mark Allen Publ. 18, pp. 1010–1013 (2009)
50. Mason, S., O'Keeffe, C., Knowles, E., Bradburn, M., Campbell, M., Coleman, P., Stride, C., O'Hara, R., Rick, J., Patterson, M.: A pragmatic quasi-experimental multi-site community intervention trial evaluating the impact of Emergency Care Practitioners in different UK health settings on patient pathways (NEECaP Trial). Emerg. Med. J. EMJ. 29, pp. 47–53 (2012)
51. Lennon, E., Kernan, R., Gallagher, J., Walshe, M., Riley, J., Keane, C., McDonald, K.: A study to evaluate the pharmacist's perspective of the heart failure patient pathway. Ir. J. Med. Sci. 186, pp. S476–S477 (2017)
52. Martin, H.M., Pedersen, L.H., Johansen, A.M.U.: Method description: user-led healthcare in patient pathways. Ugeskr. Laeger. 179, (2017)
53. Ryhanen, A.M., Rankinen, S., Tulus, K., Korvenranta, H., Leino-Kilpi, H.: Internet based patient pathway as an educational tool for breast cancer patients. Int. J. Med. Inf. 81, pp. 270–278 (2012)
54. Toronto General Hospital, University Health Network: Patient Pathways, http://pie.med.utoronto.ca/PatientTeaching/patient-teaching-projects/patient-pathways/ (Accessed: 24.05.2018)
55. Neary, P.M., Sung, R., Corrigan, M., O'Donovan, M., Cahill, R.A., Redmond, H.P.: The benefits of an interactive, individualized online patient pathway for patients undergoing minimally invasive radioguided parathyroidectomy: a prospective, double-blinded, randomized clinical trial. Surg. Innov. 17, pp. 236–241 (2010)
56. Charman, C., Al-Nuaimi, Y.: Teledermatology using National Health Service e-Referral: streamlining patient pathways. Br. J. Dermatol. 175, pp. 189–189 (2016).
57. Rothery, S., Green, H., Hale, C.: Evaluating the use of patient pathway scores in medical assessments. Emerg. Nurse J. RCN Accid. Emerg. Nurs. Assoc. 24, pp. 27–31 (2017)
58. Marschollek, M., Demirbilek, E.: Providing longitudinal health care information with the new German Health Card - a pilot system to track patient pathways. Comput. Methods Programs Biomed. 81, pp. 266–271 (2006)
59. Hallberg, S., Claeson, M., Holmström, P., Paoli, J., Wennberg Larkö, A.-M., Gonzalez, H.: Developing a simulation model for the patient pathway of cutaneous malignant melanoma. Oper. Res. Health Care. 6, pp. 23–30 (2015)
60. Balázs, T., Rakonczai, P., Frigyesy, R., Bacskai, M.: Comparison Study Based on Simulation of Patient Pathways Using Multi-State Models in Oncology. Value Health. 19, p. A380 (2016)

61. Arora, S., Cox, C., Davies, S., Kassab, E., Mahoney, P., Sharma, E., Darzi, A., Vincent, C., Sevdalis, N.: Towards the next frontier for simulation-based training: full-hospital simulation across the entire patient pathway. Ann. Surg. 260, pp. 252–258 (2014)
62. Giblett, N., Rathore, R., Carruthers, D.: Simulating the Surgical Patient Pathway for Undergraduates. J. Surg. Educ. 74, pp. 271–276 (2017)
63. Porritt, J., Baker, S.R., Marshman, Z.: A service evaluation of patient pathways and care experiences of dentally anxious adult patients. Comm. Dent. Health. 29, pp. 198–202 (2012)
64. Corless, L., Brew, I.: Patient pathway: the ideal approach. Br. J. Nurs. Mark Allen Publ. 27, pp. S14–S19 (2018)
65. Despiau, F., Bombail, M., Le Duff, G., Peoc'h, N., Labatut, R., Ceaux, C.: The patient pathway, an essential element of the care project. Soins Rev. Ref. Infirm. 28–30 (2015).
66. Manning, V., Garfield, J.B., Best, D., Berends, L., Room, R., Mugavin, J., Larner, A., Lam, T., Buykx, P., Allsop, S., Lubman, D.I.: Substance use outcomes following treatment: Findings from the Australian Patient Pathways Study. Aust. N. Z. J. Psychiatry. 51, pp. 177–189 (2017)
67. Slaich, B., Careless, A., Flynn, M.: University Hospitals of Leicester mesothelioma patient pathways April 2015 to March 2016. Poster Abstr. 16th Annu. Br. Thorac. Oncol. Group Conf. 2018. 115, Supplement 1, p. S28 (2018)
68. Tritter, J.Q., Lutfey, K., McKinlay, J.: What are tests for? The implications of stuttering steps along the US patient pathway. Soc. Sci. Med. 1982. 107, pp. 37–43 (2014)
69. Swancutt, D., Joel-Edgar, S., Allen, M., Thomas, D., Brant, H., Benger, J., Byng, R., Pinkney, J.: Not all waits are equal: an exploratory investigation of emergency care patient pathways. BMC Health Serv. Res. 17, p. 436 (2017)
70. Coughlan, J., Eatock, J., Eldabi, T.: Evaluating telemedicine: a focus on patient pathways. Int. J. Technol. Assess. Health Care. 22, pp. 136–142 (2006)
71. Cambridgeshire and Peterborough Clinical Commissioning group: Patient Pathways, https://www.cambridgeshireandpeterboroughccg.nhs.uk/health-professionals/patient-pathways/ (Accessed: 24.05.2018)
72. Whiteford, M., Byrne, P.: Talking about alcohol: communities of practice and patient pathways. J. Res. Nurs. 20, pp. 12–25 (2015)
73. Scottish Government: Patient Pathway Management, http://www.gov.scot/Publications/2007/03/19092814/4 (Accessed: 24.05.2018)
74. Jensen, H., Vedsted, P.: Exploration of the possible effect on survival of lead-time associated with implementation of cancer patient pathways among symptomatic first-time cancer patients in Denmark. Cancer Epidemiol. 49, pp. 195–201 (2017)
75. Schrijvers, G., van Hoorn, A., Huiskes, N.: The care pathway: concepts and theories: an introduction. Int. J. Integr. Care. 12, e192 (2012)
76. De Bleser, L., Depreitere, R., De Waele, K., Vanhaecht, K., Vlayen, J., Sermeus, W.: Defining pathways. J. Nurs. Manag. 14, pp. 553–563 (2006)
77. Schlieter, H., Benedict, M., Gand, K., Burwitz, M.: Towards Adaptive Pathways: Reference Architecture for Personalized Dynamic Pathways. In: 2017 IEEE 19th Conference on Business Informatics (CBI). pp. 359–368 (2017)
78. Braun, R., Schlieter, H., Burwitz, M., Esswein, W.: BPMN4CP Revised - Extending BPMN for Multiperspective Modeling of Clinical Pathways. In: Proceedings of the 49th Hawaii International Conference on System Sciences. IEEE, pp. 3249-3258 (2016).
79. Becker, J., Delfmann, P., Knackstedt, R.: Adaptive Reference Modeling: Integrating Configurative and Generic Adaptation Techniques for Information Models. In: Reference Modeling. pp. 27–58 (2007)

Understanding the Habitual Use of Wearable Activity Trackers

Annamina Rieder[1], Christiane Lehrer[1], Reinhard Jung[1]

[1] University of St. Gallen (HSG), Institute of Information Management, St. Gallen, Switzerland
{annamina.rieder,christiane.lehrer,reinhard.jung}@unisg.ch

Abstract. Given the large discrepancy between rates at which wearable activity trackers (WATs) are initially adopted and their continued use, the question concerning how sustainable use emerges arises. While IS research has found habit an important driver of sustained use, the mechanisms of habit formation have been left unexplored. To address this research gap, we conducted narrative interviews to investigate the habit-formation mechanisms behind the use of WATs. We identified two drivers of habitual WAT use and constructed five narratives that provide insights into the habit-formation processes of WAT users and possible interrupting factors. Our results provide a valuable basis for both theory and practice in explaining how sustained WAT use develops.

Keywords: Wearable, Habit, Habit Formation, Sustained Use, Continuance.

1 Introduction

The growth in the wearable industry continues as increasing sales figures and numbers of adopters of wearable activity trackers (WATs) buoy the technology's success [11, 27]. However, abandonment rates for the devices are striking [11]. A study by Endeavour Partners reported that users' minimum attrition rate is as high as 30 percent within the first six months [18]. However, sustained use is a prerequisite for inducing long-term changes in health behavior, which is most WAT devices' key value proposition. Information systems (IS) research has identified habit as a key driver of continuous IS use [e.g., 20, 3], but most of its studies of habit and post-adoption phenomena are quantitative in nature and focus on the relationship between habit and factors like continuance, leaving the underlying mechanisms of habitual IS use unexplained. Research that has investigated WATs [e.g., 1, 12] has provided only a few insights into users' perspectives and experiences. Clarifying how habitual use patterns emerge in the context of WATs and post-adoption phenomena would help in designing WATs in a way that supports users' sustainable behaviors and habits. Against this background, our research question asks, *What are the underlying mechanisms of habitual WAT use?*

We used a narrative interview technique in qualitative interviews with ten habitual WAT users to gain a longitudinal insight into the users' experiences with the devices. Eyal's (2014) IS habit formation model, which is grounded in psychology and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

neuroscience research, provided a valuable theoretical lens through which to investigate the cognitive and behavioral processes of habitual WATs users and the factors that can interrupt habitual use.

The paper proceeds as follows. First, we discuss the theoretical background on wearables, the role of habit in IS, and the mechanisms behind habit formation. Then we present our methodology and the results of our analysis. Finally, we conclude with a discussion of our results and limitations and a proposal for further research.

2 Theoretical Background

2.1 Wearable Activity Trackers

"Wearables" refers to any electronic computing device worn on the body and that uses sensor technology so users can track their personal activities and vital parameters [32, 29]. Even though a wide range of devices (e.g., wristbands, glasses, clothing) fall under the umbrella term of wearables, we focus on wearable devices that track physical activity: WATs. Wearable devices that are designed only to track other health-related parameters, such as fertility, blood pressure, or blood glucose levels, do not fall into the category that we investigate in this paper. WATs, which usually take the form of bracelets and watches, are the most popular category of the wearables market [11]. These devices allow users to self-optimize by providing them with insights into their physical performance through self-tracking of such parameters as steps, sleep, sports activities, and food consumption [29, 10]. Users can usually interact with WATs directly over the device's screen or via mobile applications, which are typically accessible using a smartphone.

Because of the possible positive effects of WATs on personal health, the healthcare sector places considerable value on their potential. The intended behavioral outcomes of wearable use, such as increased activity levels and conscious nutrition, are key preventive measures that can improve public health and reduce strain on healthcare providers and insurers [10]. However, this promising outlook is dampened by the large discrepancy between adoption rates and sustained use. Findings from the US show that, despite adoption rates higher than 20 percent in 2014 and estimated adoption rates higher than 35 percent by 2020, attrition rates are as high as 30 percent within the first six months of use [18, 27]. Consequently, many WAT users do not reach the phase of sustained long-term use, which is a better indicator than mere adoption with which to assess the success of WATs [4]. Research has identified several issues that lead to discontinuance, including short battery life, inaccurate sensors, privacy concerns, insufficient data, and irrelevant information provided to users [6, 7, 27]. Some authors have also suggested that the design of WATs insufficiently incorporates findings from behavioral theory [21, 6, 7].

To date, research on wearables has predominantly followed a quantitative approach to examine the effectiveness of using WATs in health behavior interventions [e.g., 12], and only a small fraction of articles has focused on the experiential side of WATs and the qualitative analysis of the user's perspective [e.g., 2, 13, 24]. The majority of articles has addressed only the adoption process, so we lack qualitative experience-centric research that explores sustained WAT use in the post-adoption phase.

2.2 Habit in Information Systems Research

Research on continued IS use in the post-adoption phase has consistently highlighted the importance of habit as a key predictor of continued IS use [20, 3, 14, 22, 19]. Habit can be defined as the automatic or unthinking performance of specific behavioral sequences, such as the use of an IS artifact, to obtain a certain goal in response to environmental cues [20, 3]. Automation occurs through the processes of repetition and learning.

Studies that have investigated the effect of habit on continued IS use have shown two primary influences: a direct positive influence on IS continuance, continuance intention and satisfaction [3], and an indirect suppressing effect that moderates the relationship between continuance intention and actual continuance [20]. Bhattacherjee and Lin's (2015) findings also supported this link, corroborating the notion that automatic behavior reduces the need for intention. Limayem et al. (2007) identified four antecedents of habits in IS use: frequency of the behavior, satisfactory outcome, a stable context, and the use of various features of the IS in question (comprehensiveness of use). Building on these findings, a handful of studies have investigated continued IS use and the role of habit in specific use contexts, including participation in online communities [14], online gambling [22], mobile location-based services [19], and quitting Facebook use [31].

Whereas the extant research on habit in IS use has involved variance theory, predicting the relationships between antecedents and outcomes, Kim (2009) offered a process view on IS continuance and the influence of habit. Drawing on a process model from cognitive psychology that outlines how memory is formed, Kim sheds light on the process that drives post-adoption behaviors. However, the model does not provide a detailed explanation of the process of habit formation itself.

In summary, the extant IS research assigns habit an important role in predicting IS continuance, but it has yet yielded only a few attempts to untangle the mechanisms and processes behind habit formation. Therefore, we turn to Eyal (2014), who developed a process model to explain habit formation with digital products and services.

2.3 Mechanisms of IS Habit Formation

Process of IS Habit Formation. Eyal's (2014) process model indicates that IS habit formation takes place in four-step cycles: trigger, action, variable reward, and investment (Figure 1). This view is in line with the definition of habit from social psychology, which understands habit as automated dispositions to respond to specific cues to obtain certain goals or end-states [25, 33]. Responses are learned sequences of acts, memorized through a process of repetition that have led to satisfactory outcomes in the past [25]. Habit formation takes place when a trigger—that is, an environmental cue—instructs users in a more or less explicit way about what action to take next [33, 25]. If users are sufficiently motivated and capable of performing the behavior, they will execute the action [9]. The action always takes place in anticipation of a reward, which users expect to receive upon completion of the action [25]. After multiple repetitions, users start building associative links between the trigger and the action and between the action and the reward, which lead them to repeatedly and effortlessly perform the behavior when they are confronted with a similar context [28]. Users also typically "invest" in a specific technology by putting something of value into the system (e.g., effort, time, skill acquisitions) [8]. Along with the processes of associative learning, such investments increase the likelihood that users will repeatedly pass through the process [26]. Repeating this process with high frequency is critical to habit formation [17, 25].

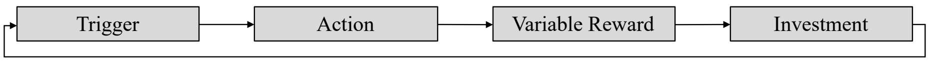


Figure 1. IS habit formation model (based on [8])

Trigger. A trigger can be understood as an "actuator of behavior" [8, p. 7], the first step in the habit cycle. Triggers are goal-related in so far as they cause individuals to desire a certain reward that they expect on execution of a specific behavior [33, 8]. The literature differentiates between external and internal triggers [9, 33] such that external triggers are stimuli sent out by the technology that tells the user what behavior to perform next (e.g., a push notification that reminds the user to take a specific number of steps every hour), while internal triggers do not require a sensory stimulus but are firmly installed in users' minds and automatically activated as certain thoughts, emotions, or routines come up. Examples of internal triggers are feelings of boredom or hunger, certain weekdays, and preceding actions. Having laced up one's running shoes could serve as an internal, routine-based trigger for users to activate their WATs' sports monitoring function.

Action. Action refers to the behavior performed in response to the trigger and in anticipation of the reward [33]. Performing the action requires the user to be sufficiently motivated and able to do so [9]. Pursuing the goals or end states that result from performing the action can include seeking to achieve a positive goal but also avoiding a negative one. Ability, that is, the ease with which a user can perform an action, also influences the likelihood that the action will occur. Ability involves the user's personal capabilities and skillset, as well as contextual factors like time restrictions and the physical environment [25]. In the context of WATs, an action could include making the missing number of steps to fulfill the daily goal. Thereby, the user's motivation to execute this action could, for example, be negatively influenced by the weather, while restrictions to the user's ability could include not having sufficient time or being temporarily incapacitated due to an injury.

Variable Reward. The successful completion of the action is followed by a variable reward, a "positive reinforcer" of behavior [28, p. 89; 33]. With repeated encounters with a similar situational context, associative links between the context and response that have yielded a specific reward are established [33, 25]. Thus, what drives the user to comply with the trigger is not the reward itself but its anticipation, which causes the emission of neurotransmitters in the human brain [28]. The reward should be variable in so far as it provides the user with some degree of novelty [8]. Rewards can take a variety of forms, including social rewards, rewards that fulfill the human desire to acquire resources and information, and rewards that have intrinsic benefits. Eyal (2014) highlighted the importance of the fit among the reasons an individual uses a product, his or her internal triggers and motivations, and the reward. Related to WATs, rewards could, for example, include receiving badges, seeing an upward tendency in performance, or improvements in wellbeing.

Investment. Users' investments in a product or service (e.g., time, data, effort, social capital) influence the value they assign to it [cf. 26]. In the case of technology and software, such investments can enhance the users' experience and, thus, increase the likelihood of re-engagement [8]. Stores of value with technology can take the form of data, followers, acquired skill, content, or reputation. In the context of WATs, investments could comprise a personal training history or a network of friends and followers.

3 Methodology

Since empirical evidence of the process through which habitual WAT use is formed is scarce, we employed an inductive, qualitative approach using narrative interviews, following established principles [16, 5], to capture individual users' experiences and a longitudinal picture of the individual's use history. In this interview technique, the interviewee recalls and gives an account of a past event [16]. Typically, the interview's organization and structure are left to the interviewee, and the interviewer does not intervene with the narrative until the interviewee has recognizably ended his or her recitation. Thus, this interview technique can help to overcome common biases like social desirability, as well as patterns of interaction in the interview, issues related to the wording and placement of questions, and topics and terminology brought in by the interviewer [16].

3.1 Data Collection

We interviewed ten WAT users who were based in Switzerland, which, with a 7.6 percent market penetration in wearable devices in 2017, is one of the most advanced markets for wearables in Europe [30]. In line with the narrative interview technique, we used no interview guideline [16]; instead, we used a pre-formulated initial stimulus to ignite the interviewees' narratives: *"I would like you to tell me the story of your activity tracker. This means from the moment you got it, until today."* We did not intervene until the interviewees had recognizably finished their narratives. Then we took up topics the interviewees had mentioned in their initial narratives to trigger additional accounts. All of the interviews were held in person, and interviewees were ensured anonymity and confidentiality in the statements they made during the interview. The narratives' duration (ranging from 24 to 68 minutes, with an average length of 42 minutes) and level of elaboration varied widely. We recorded and transcribed the interviews verbatim to ensure rigorous and transparent analysis of the resulting data. Interviews and transcriptions were done in the participants' native language, either German or Swiss German dialect, and were processed by native German speakers. The quotations presented in this paper were translated into English.

As we sought depth over breadth of information, a sample of ten participants was adequate to gain rich insights into personal use histories and to identify various habit formation processes. We followed a criterion-based sampling strategy [23], taking the respondents' gender, age, and duration of use into account. Although generalizability is not a primary objective in qualitative research, we tried to distribute interviewees evenly across both genders and to cover a broad age range. Our sample consists of four females and six males whose ages range from 18 to 58 years (M=33.5), and it includes both students and professionals (Table 1). The sample slightly over-represents younger people, but younger people tend to adopt WATs more frequently than older people do [30]. We also chose continuous WAT users (i.e., those who used their devices past the trial phase) who had used their WATs for at least six months, although a few had used them for several years. The use frequency varied across interviewees and largely depended on the motivation for use. Most interviewees used the device every day and even at night, while others only wore it only when they engaged in sports. Most of them used their WATs to track their daily activity, support their sports performance, and improve their health. Some reported using the devices to explore the novel technology out of curiosity. The four functionalities used frequently and long-term by more than three participants were step-counting (nine out of ten participants), sports-activity monitoring (6 participants), competition (5 participants), and heart-rate feedback (3 participants). Fewer than three participants used nutrition tracking and sleep tracking habitually, which others used mainly on a trial basis.

Table 1. Sample characteristics

Alias	*Age*	*Gender*	*Occupation*	*Total use duration*	*# WATs used*
INT1	25	F	Student	> 24 months	2
INT2	24	F	Student	6 - 12 months	1
INT3	57	F	Project manager	12 - 24 months	2
INT4	58	M	Consultant	> 24 months	4
INT5	41	M	Entrepreneur	12 - 24 months	2
INT6	41	M	Entrepreneur	6 - 12 months	1
INT7	25	M	Student	6 - 12 months	1
INT8	20	F	Student	6 - 12 months	1
INT9	18	M	Student	12 - 24 months	2
INT10	26	M	Bartender	6 - 12 months	1

3.2 Data Analysis

The data analysis process broadly followed Chatman's (1978) recommendations. In the first step, we used narrative coding to decompose the narrative into its elements so we could extract the kernel sequence of events. According to Chatman (1978), the story of a narrative consists of *events* and *existents*. Events can be understood as changes of state that are driven by *actions* executed by significant agents (e.g., *"In the beginning, I measured my heart rate relatively consistently" (INT4)*;*"This was when I actually stood up and started walking through the apartment"* (INT1)) and *happenings* that occur but are not directly influenced by the agents (e.g., *"This fitness tracker brought to my attention what I was eating and how often I was exercising"* (INT2); *"The problem was that it just wasn't working most of the time. Then there was something* *wrong with the software, then this, then that"* (INT3)). Existents, on the other hand, are descriptive elements of the story that are constituted of *characters*, which are animate or inanimate agents that advance the plot (e.g., *"I wasn't really a sports enthusiast"* (INT9); *"My fitness tracker is a first-generation Apple Watch"* (INT5); *"The platform has more analytical functionalities, where my running coach could see my progress"* (INT4)) and the *setting*, which is the space in which the characters exist and act (e.g., *"I had two, three colleagues in school who had this too"* (INT9); *"During the semester break we took off the device when we had to study because we wouldn't even make 1,000 steps per day"* (INT1)).

We differentiated between kernel and satellite sequences of events. Kernel events are actions and happenings that advance the plot and are central to the participants' narrative. For example, kernel events were identified if the participants gave an in-depth and lengthy recounting of their use history of a key WAT functionality. Satellite events are not central to the plot, so omitting them does not alter the story. For example, satellite events include when participants mention a WAT functionality only briefly or tell how they used it only on a trial basis.

In the second step, in which we focused on the kernel sequences of events, we identified the WAT functionalities that the participants used habitually: sports-activity monitoring, heart-rate feedback, step-counting, and competitions. Indicators of habitual use patterns were long duration of use, regular use, and signs of automaticity (e.g., when a user mentioned that he or she always and unthinkingly performed certain actions). For those functionalities, we identified the underlying drivers of habitual use. While our research process was exploratory, we were sensitized by the concepts of Eyal's (2014) IS habit-formation model and used the model to identify how the steps of trigger, action, reward and investment were represented in the data along the narratives. The IS habit-formation model proved to be an appropriate theoretical lens, as the interviewees went through steps of the process when they used their devices' various functionalities.

4 Results

Our analysis revealed two drivers of habitual WAT use identifiable across narratives. On the one hand, habitual WAT use was driven by users' offline habits performed independently of WAT use, as it was the case for the functionalities sports activity monitoring and heart rate feedback. On the other hand, the technology itself fostered habitual WAT use by sending out triggers, as with step counting and competitions. Each of the following subchapters covers one of the two motors. The first subchapter illustrates habitual WAT use driven by offline habits by presenting the composed narratives for the functionalities sport-activity monitoring and heart-rate feedback. The narratives in the second subchapter on the motor habitual WAT use driven by the device, are related to the functionalities step reminders, activity statistics, and competitions. The narratives presented below follow the logic of the IS habit formation model, comprising the three impelling elements *trigger*, *action*, *reward*, as well as potential *interrupting factors* to habitual use found in the data. We purposely left out the *investment* element of the model, as we did not find any explicit mentions in relation to specific WAT functionalities.

4.1 Habitual WAT Use Driven by Offline Habits

Habitual Use of Sport Activity Monitoring

Sport activity monitoring allows users to track their exercise (mainly running) using performance indicators like distance, pace, and blood oxygen consumption. A characteristic of this functionality is the need to switch it on and off to monitor the activity while it is being performed. The interviewees who habitually used sport activity monitoring typically had already established a sports habit—usually running—or wanted to develop one before they adopted a WAT. Most had a clear focus on sports and intended to support their sporting goals using a WAT or were interested in monitoring and the data as such.

It's a Christmas gift that I asked for because I went running regularly. I basically used the tracker for running. I also wore it during leisure time and at work, but I mainly used the functionalities for running, like to measure the distance I had run. It also monitored the heart rate, which was useful sometimes, and especially the time per kilometer, which was very important to me. I w0anted to see the progress I made. (INT10)

Sport activity monitoring was in all cases cued by a pre-existing (offline) sports habit, which can be classified as an internal trigger. Intending to or preparing to engage in sports would automatically trigger the participants' subsequent action of switching on the sports tracking mode. In addition, several interviewees engaged in information retrieval after the sports training sessions and even processed the information further by, for example, sharing it with their social networks or feeding it into other data analytics platforms.

When I bought the Suunto, I started trying out the apps. I found that very exciting. There are lots of interfaces through which you can export data. My main platform is Strava, which allows me to see my activities. For example, I went skiing, and you can see where I took photos. I'm sure there are some people who would call this profile neurosis, but I think it's fun. (INT4)

Using sport activity monitoring brought the interviewees multiple kinds of rewards that were all based on the technology. The interviewees often mentioned that the statistics and history motivated them and allowed them to monitor their progress and path toward reaching their goals, which had a confirming effect. Moreover, quantification and visualization of the sports training sessions were considered rewarding in themselves because they led to insights into novel parameters. Monitoring the training sessions also gave the interviewees a sense of achievement because parameters like calories burned, time trained, and distance covered were made transparent. Sports activity monitoring also rewarded them with social recognition when they shared their achievements through social network sites and received positive responses. In addition, spending time with the monitoring data was perceived as entertaining, as they could, for example, keep track of their routes on Google Earth and add the photos they had taken during the training.

Of course, it is great at the beginning, when you're proud or when you do your ten kilometers for the first time and pass your limits for the first time, or when you do your first trail run. Of course, you do want to share this and document all of it. (INT5)

The most common cause for interruptions of the habitual use of sport activity monitoring was loss of the reward. Besides not being able to look at the statistics because of technical problems or a general lack of interest in seeing one's progress, interviewees described how the documentation lost its novelty and excitement after a time. Other such interrupting factors were changes in underlying offline habits that resulted in loss of the trigger and broken devices.

For me, it was the case that it didn't particularly motivate me anymore. It was more like you've been outside for an hour in fresh air and had a good time, so documentation wasn't that important to me anymore because I realized I felt good. You end up doing it regularly – when you move from compulsion to enjoyment. That makes the difference. (INT5)

Habitual Use of Heart-Rate Feedback

Heart-rate feedback, which provides users with close to real-time information about their heart rate, is constantly running in the background. This functionality was mainly used habitually by interviewees who also reported strong, established sports habits or wanted to develop them before adopting a WAT. Using heart-rate feedback served specific goals: improvements in distance and pace or high calorie burn.

The story of how I came across the fitness tracker was that I started to go running on a regular basis, like two to three times a week, and I had to measure my heart rate because it was not supposed to exceed 160. I think my personal trainer was afraid I would have a heart attack. (INT5)

A pre-existing (offline) sports habit serves as an internal trigger, so engaging in a habitual activity like running or doing cardio exercise automatically set the cycle in motion.

I also used the heart-rate monitor during sports. I always checked it; for example, when I went running, I would eventually have to make a cardio peak and then I always checked whether it was over 160. (INT2)

The interviewees report two subsequent actions in the case of heart-rate monitoring: constant interaction with the device during the activity to check the heart rate and immediate adjustment (e.g., breathing, pace) when the heart rate deviated from the target.

I tried to keep a lower pulse on longer distances and not go at full intensity. I also tried to slow down a bit when my pulse was at 160 or so. The tracker was really useful for when you want to improve your running performance. Sometimes, even when you don't have the impression you're going so fast, you take a look at your pulse and the pace and you say, "Oh, 170 is maybe a bit too high for a flat route—should probably slow down." (INT10)

The rewards interviewees got from using heart-rate monitoring were not usually delivered by the technology but were internal. Interviewees reported rewards that were personal in nature, such as feeling good, improving, being able to run longer in a healthier way, and getting to know their bodies.

However, through repetition and learning, interviewees felt they had acquired a subjective feeling for their bodies and heart rates, which obviated the need to use the digital heart-rate feedback provided by WATs. Changes in underlying habits that resulted in loss of the trigger, and broken devices are other factors that interrupted the habitual use of the heart-rate feedback.

Once you've reached a certain level–this was the case with me after some three months, and after half a year it got really easy–where you can run for an hour plus, you've learned to deal with it and you also feel for *yourself which pulse is good for you, then you don't need the technical support anymore to guide you. As you get to know your body better, the device becomes less informative.* (INT5)

4.2 Habitual WAT Use Driven by the Device

Habitual Use of Step Reminders

The step-counter of a WAT is always running in the background, recording each movement that is sensed as a step. Most WATs come with pre-set hourly and daily step goals or let users specify goals themselves. External triggers in the form of push notifications on the WAT screen or the user's smart phone drove the habitual use of the step-counting functionality for our interviewees. The devices reminded users to move if they were to achieve their hourly activity goals or informed them of how many steps were left to reach their daily step goals. Such push notifications triggered the interviewees to interact with their WATs and to engage in offline actions—that is, to take the number of steps that were necessary to reach a certain goal.

I always got something five or ten minutes before the hour saying, "Hey, 300 steps left for the hourly goal" or so. It was vibrating, and then you realized, "Oh, I've been sitting around for an hour looking at the computer. I'll get up and go to the toilet or somewhere else just to walk a few meters." You're always reminded when an hour passes and another hour passes, and you haven't moved at all. (INT7)

Interviewees reported various rewards upon executing the triggered behavior. Complying with the cue usually left them with the feeling of being energized and more fit during their daily activities. Moreover, as making hourly step goals also added to the overall step count, they felt that compliance contributed to overall goal achievement. With WATs that were using a different technique like a punishment logic with a red bar that appeared on screen after a certain amount of sedentary time, executing the action was perceived as a relief.

The more bars you had, the longer it took to get rid of them, so if you had one bar, you could cross your apartment two or three times before it eventually disappeared. But when you had five bars, this meant you would have to move for five or ten minutes to get rid of all of them. Yes, the red bar really annoyed me. (INT8)

The most common interruption to the habitual use of step reminders the interviewees mentioned was the loss of the trigger if users stopped wearing the WAT regularly. Even though the interviewees intended to maintain their activity levels, they reported being unaware of their hourly number of steps and so were not cued to move.

I only stand up when I really have to. If I am not made aware of it, I don't think, "Oh, I haven't gotten up for three-quarters of an hour straight." (INT9)

Habitual Use of Activity Statistics

WATs provided the interviewees with a wide variety of statistics and historical data about their physical activity, graphically represented in dashboards that were accessible via smartphone app. Unlike using step reminders, using these functionalities requires the user to access and retrieve information actively.

I was curious to see how many steps I take and how this changes over the course of the week–how this changes based on whether you're doing sports or are just at university. Even if I did not reach my step goal each day, I just concentrated on the weekly average and made sure that I got a daily average of 10,000 steps per day, and if I succeeded in that, I didn't care too much when I had a day on which I had fewer steps and another one with more. Then it was kind of balanced. (INT1)

For some functionalities, the triggers are external, but they take a passive form, as they remind users of the goals they have set and to interact with the technology when it catches their attention. For example, merely seeing the wristband on one's wrist or seeing the WAT app on the screen of a smartphone app can serve as a trigger.

Triggered actions can be comprised of only interaction with software components or can also include execution of an offline behavior. Interviewees regularly checked their numbers of steps, consulted their statistics and histories, and looked into the progress they had made toward reaching their goals. However, most interviewees used their WATs' functionalities not only by interacting with them but also by adapting their offline behavior based on their progress toward their goals. Most participants engaged in extra activity if they had not reached their step goal yet.

When you're looking at your device and you see that you have half the steps that you normally do, it's a bit shocking and you think you'll have to do something about it. (INT7)

Interviewees received a variety of rewards from the technology for executing the actions it triggered, including a sense of achievement, confirmation from an upward trend over the use history, and avoidance of "punishment" when the interviewee's goal was to prevent a downward movement in data. Engaging in offline behavior in response to the activity history also added to interviewees' well-being and their perceptions of their fitness, as it kept them active and moving.

When you went running, you quickly got 15,000, 16,000 steps, and that was great. I had days on which I had 25,000 steps–that made me proud. And then it also told you the calories, and then you saw you really burned a lot. (INT3)

For some users, learning effects decreased the interviewees' perceptions of rewards and their excitement from using the activity history functionality for a period of time.

It eventually loses its novelty value or you eventually know and get a feeling for how many steps you've walked. (INT6)

The interviewees' ability to perform the offline behavior was restrained in some cases when they faced changing schedules or lost awareness of the number of steps they had taken after discontinuing use of their WATs.

I know that I don't do a lot during the exam period, and I think that I can't because I have tons of other stuff to do. But when you have it in black and white that you have 1,000 steps over the whole day, you just don't wear the WAT at all to silence your conscience.(INT1)

Habitual Use of Competitions

WATs offer users the opportunity to compete against other users through a social network. The functionality, which is based on step-counting, lets users initiate or accept competitions and provides them with reports in the form of rankings or reminders entailing the status of the competition. Competing against other WAT users was rarely the interviewees' initial motivation to use a WAT; instead, they learned about the function when they adopted their devices and tried it out with their friends. The strength of the habit of using competitions was affected by one's social network's motivation and involvement.

In the beginning, it was fun to see how many steps I take, but then two or three colleagues at school had the device too and there was the possibility to engage in challenges against each other. I eventually started walking home from school just to win the challenge. (INT9)

Engagement in competitions was triggered external to the device when the interviewees were invited to compete against other users or initiated the competition themselves. To participate in a competition, interviewees had to take the first action of initiating or accepting an invitation. Further actions included checking the competition status and the rankings in the social network. The interviewees also often engaged in an offline behavior like taking more steps to win a competition.

It motivated me. I purposely took the stairs or, when I had to take the subway for one stop, I walked instead because I desperately wanted to beat my boyfriend. (INT2)

Rewards for participating in a competition were usually delivered by the technology. Interviewees drew social recognition from competitions, as others got to see how active and fit they were and offered admiration. Interviewees also received confirmation through winning a competition that confirmed that they were better or more active and sporty than their competitors, which served as a form of reward. Besides being entertaining, competitions also supported interviewees' fitness by prompting them to take even more steps.

I believe this is what motivates people most–comparing themselves to others. You look up on Freeletics or Linkedin what all the other people are doing and you think, "Ah, shit, I totally don't have my life under control," and then you do something about it. (INT2)

The reason for ending the habit was usually that interviewees lost interest in the rewards. Similar to step-counting, some interviewees lost interest in the function after the excitement and novelty value had vanished and rewards had become predictable.

At some point, after all these competitions were nothing special anymore, it just wasn't an advantage to me anymore. (INT2)

Other problems had to do with restricted social networks and frequently losing competitions, both of which negatively affected the rewards connected with the behavior.

With the one colleague I used it with, I didn't have any chance. Even when I thought I had done lots of steps, like 40,000 or so a day, he ended up getting 60,000. Then I thought I'd have to stop this, as it started annoying me. (INT1)

5 Discussion

This study investigates the mechanisms behind habitual WAT use. Based on narrative interviews, we identified two distinctive drivers of habitual WAT use and constructed five narratives that provide insights into WAT users' habit-formation processes and possible interrupting factors. The two drivers of habitual WAT use are pre-existing offline habits and the device itself. When offline habits drive WAT use, WAT use is embedded into existing behavioral sequences that have become habitual prior to WAT adoption. When use habits are driven by the device, users are incited to learn and perform new behaviors.

Our research contributes to the literature on habit and wearables and has implications for theory and practice. First, we contribute to the IS literature on habit [e.g., 20, 3] by unveiling the process by which use habits are formed in the context of WATs. Two drivers of habitual WAT use are pre-existing offline habits and the technology itself, but only one of the two drivers tends to be in effect with a particular user—that is, users either habitually used their WATs in response to an offline habit (INT3, 4, 5, 10) or were driven by the technology (INT1, 6, 7, 8, 9), rather than being motivated by both drivers (INT2). Moreover, taking into consideration Limayem et al.'s (2007) research, which provided insights into antecedent conditions of IS habit, our findings indicate that comprehensiveness of use is not a substantial driver of habit in the WAT context or for our sample. On the contrary, we observed a tendency toward long-term use and stronger use habits the more users were focused on only a few functionalities of the WAT. Wide adoption of WAT features was often associated with trial usage or weak habits.

Second, we contribute to research on wearables by offering insights into the experiential side of WATs, particularly by providing information about habit-formation processes and habitual use of WATs. Unlike to several other kinds of IS, WAT use cannot be conceptualized simply as interaction with the device's software components, as WATs aim to induce offline behavior, and the performance of such behavior constitutes a qualified form of WAT use. Our findings suggest that a much of what leads to interruption of sustained WAT use resides with the user, although the perspective of current WAT research predominantly allocates interrupting factors to the technological sphere.

Our research has several implications for theory and practice. First, our research suggests that it is reasonable to conceptualize habitual WAT use along its functionalities because users develop use habits for some functionalities, but not for others. Second, Eyal's (2014) IS habit-formation model is useful in explaining IS habits on a fine-grained level, as we found evidence for the stages of the model in our data. Therefore, IS researchers could apply the model to examine similar IS habit-formation processes. Third, our research offers producers clarity about the habit-formation processes that are related to WAT use and about potential interrupting factors. Building on such information, WATs designers can design them in a way that supports particular stages of the habit-formation process and that anticipates and counteracts the interrupting factors that are obstacles to long-term WAT use. Fourth, as habitual WAT use is driven in part by pre-existing offline habits, producers must shift their focus away from technology-driven behaviors to consider users' existing behavioral patterns and habits.

6 Limitations and Further Research

The present research is subject to some limitations. Because this research is qualitative, we limited the number of interviewees we used for analysis, so we may not have heard all variations of use cases in the WAT context or have covered all functionalities equally. Future research may consider interviewing users with other use cases and contexts. Although the narrative interview technique is especially valuable when it comes to exploring longitudinal, process-like issues with highly individual character, some themes could be under-represented in the data, which could be why we found no evidence of the investment element in our data. While interviewees mentioned investment-related aspects of WAT use on a general level, such as appreciating the historic data or having gained specific use skills, we were unable to relate them to specific functionalities. Future research could verify this finding and investigate the role of investment in the habit-formation process with WATs. Moreover, habits always underlie automated cognitive processes that users do not necessarily consciously perceive. Therefore, when we relied on the interviewees' memories, we had to use interpretive proxies to identify such processes. However, also other forms of qualitative studies, as well as quantitative studies, address this particular limitation, so future research could use other research methods (e.g., mining WAT use data) to triangulate and establish reliability. Finally, future research could go a step farther to look into behavior changes that result from sustained WAT use and assess how WATs are keeping up with their goal of inducing behavior changes in their users.

References

1. Barwais, F.A., Cuddihy, T.F., Tomson, L.M.: Physical Activity, Sedentary Behavior and Total Wellness Changes Among Sedentary Adults: A 4-Week Randomized Controlled Trial. International Journal of Environmental Research and Public Health 12(1), 414-427 (2013)
2. Becker, M., Kolbeck, A., Matt, C., Hess, T.: Understanding the Continuous Use of Fitness Trackers: A Thematic Analysis. PACIS 2017 Proceedings (2017)
3. Bhattacherjee, A., Lin, C.-P.: A Unified Model of IT Continuance: Three Complementary Perspectives and Crossover Effects. European Journal of Information Systems 24(4), 364-373 (2015)
4. Buchwald, A., Urbach, N., Letner, A., von Entress-Fuersteneck, M.: Towards Explaining the Use of Self-Tracking Devices: Conceptual Development of a Continuance and Discontinuance Model. ICIS 2015 Proceedings (2015)
5. Chatman, S.: Story and Discourse: Narrative Structure in Fiction and Film. Cornell University Press, Ithaca (1978)
6. Chiauzzi, E., Rodarte, C., DasMahapatra, P.: Patient-centered Activity Monitoring in the Self-management of Chronic Health Conditions. BMC Medicine 13(77), 1-6 (2015)
7. Cleland, I., Nugent, C., Lee, S.: The Ground Truth Is out there: Challenges with Using Pervasive Technologies for Behavior Change. EAI International Conference on Pervasive Computing Technologies for Healthcare 2016 Proceedings (2016)
8. Eyal, N.: Hooked: How to Build Habit-Forming Products. Penguin, New York, NY (2014)
9. Fogg, B.J.: A Behavior Model for Persuasive Design. International Conference on Persuasive Technology 2009 Proceedings (2009)
10. Gimpel, H., Nissen, M., Görlitz, R.A.: Quantifying the Quantified Self: A Study on the Motivation of Patients to Track Their Own Health. ICIS 2013 Proceedings (2013)
11. IDC: Global Wearable Market Grows 7.7% in 4Q17 and 10.3% in 2017 as Apple Seizes the Leader Position, Says IDC. (2018, 1 March)
12. Jakicic, J.M., Davis, K.K., Rogers, R.J., King, W.C., Marcus, M.D., Helsel, D., Rickman, A.D., Wahed, A.S., Belle, S.H.: Effect of Wearable Technology Combined With a Lifestyle Intervention on Long-term Weight Loss: The IDEA Randomized Clinical Trial. Journal of the American Medical Association 316(11), 1161-1171 (2016)
13. Kari, T., Koivunen, S., Frank, L., Makkonen, M., Moilanen, P.: Critical Experiences During the Implementation of a Self-Tracking Technology. PACIS 2016 Proceedings (2016)

14. Khansa, L., Ma, X., Liginlal, D., Kim, S.S.: Understanding Members' Active Participation in Online Question-and-Answer Communities: A Theory and Empirical Analysis. Journal of Management Information Systems 32(2), 162-203 (2015)
15. Kim, S.S.: The Integrative Framework of Technology Use: An Extension and Test. Management Information Systems Quarterly 33(3), 513-537 (2009)
16. Küsters, I.: Narrative Interviews: Grundlagen und Anwendungen (2nd Ed.). VS Verlag für Sozialwissenschaften, Wiesbaden (2009)
17. Lally, P, van Jaarsveld, C.H.M., Potss, H.W.W., Wardle, J.: How Are Habits Formed: Modelling Habit Formation in the Real World. European Journal of Social Psychology 40, 998–1009 (2010)
18. Ledger D. & McCaffrey D.: Inside Wearables: How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement (2014, January)
19. Lehrer, C., Constantiou, J., Hess, T.: Examining the Determinants of Mobile Location-Based Services' Continuance. ICIS 2011 Proceedings (2011)
20. Limayem, M., Hirt, S.G., Cheung, C.M.K.: How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance. Management Information Systems Quarterly 31(4), 705-737 (2007)
21. Lyons E.J., Lewis Z.H., Mayrsohn B.G., Rowland J.L.: Behavior Change Techniques Implemented in Electronic Lifestyle Activity Monitors: A Systematic Content Analysis. Journal of Medical Internet Research 16(8), 1-15 (2014)
22. Ma, X., Kim, S.H., Kim, S.S.: Online Gambling Behavior: The Impacts of Cumulative Outcomes, Recent Outcomes, and Prior Use. Information Systems Research 25(3), 511-527 (2014)
23. Miles, M.B. & Huberman, A.M.: Qualitative Data Analysis: An Expanded Sourcebook (2nd Edition). Sage, Thousand Oaks (1994)
24. Miyamoto, S.W., Henderson, S., Young, H.M., Pande, A., Han, J.J.: Tracking Health Data Is Not Enough: A Qualitative Exploration of the Role of Healthcare Partnerships and mHealth Technology to Promote Physical Activity and to Sustain Behavior Change. JMIR mHealth uHealth 4(1), 1-12 (2016)
25. Neal, T., Wood, W., Quinn, J.M.: Habits – A Repeat Performance. Current Directions in Psychological Science 15 (4), 198-202 (2006)
26. Norton, M.I., Mochon, D., Ariely, D.: The IKEA Effect: When Labor Leads to Love. Journal of Consumer Psychology 22(3), 453-460 (2012)
27. Pemberton Levy, H.: Wearable Technology Beyond Smartwatches (2015, October 21)
28. Schultz, W.: Behavioral Theories and the Neurophysiology of Reward. Annual Review of Psychology 57, 87-115 (2006)
29. Sjöklint, M., Constantiou, I., Trier, M.: The Complexities of Self-Tracking - An Inquiry into User Reactions and Goal Attainment. ECIS 2015 Proceedings (2015)
30. Statista, https://de.statista.com/outlook/319/155/wearables/schweiz (Accessed: 22.08.2018)
31. Turel, O.: Quitting the Use of a Habitual Hedonic Information System: A Theoretical Model and Empirical Examination of Facebook Users. European Journal of Information Systems 24(4), 431-446 (2015)
32. Trickler, C.: An Overview of Self-Monitoring-Systems. SAIS 2013 Proceedings (2013)
33. Verplanken, B., Aarts, H.: Habit, Attitude, and Planned Behaviour: Is Habit an Empty Construct or an Interesting Case of Goal-directed Automaticity? European Review of Social Psychology 10(1), 101-134 (1999)

On the Fit in Fitness Apps: Studying the Interaction of Motivational Affordances and Users' Goal Orientations in Affecting the Benefits Gained

Robert Rockmann[1] and Christian Maier[2]

[1] Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany
robert.rockmann@hs-neu-ulm.de
[2] University of Bamberg, Bamberg, Germany
christian.maier@uni-bamberg.de

Abstract. Lacking regular physical activity is a pertaining problem in most western societies. Fitness apps are positioned to address this issue by offering motivational affordances to the user, which aim to enhance motivation and increase physical activity: self-monitoring, rewards, and social comparison. Yet research provides inconclusive results about their effectiveness. For clarification, this paper draws upon Achievement Goal Theory and theorizes how and why motivational affordances vary in dependence of users' motivation-relevant goals in supporting motivation and physical activity. Empirical validation among 283 fitness app users generally supports that motivational affordances need to be congruent with users' underlying goal orientations to achieve the benefits. As such, this paper contributes to fitness app research by resolving prior inconsistencies, offers a theorizing on motivational affordances and individual motivation-relevant differences, and aids practice in designing fitness apps.

Keywords: Fitness apps, Motivational affordances, Achievement goal theory.

1 Introduction

January, 12th is unofficially called the *Quitters Day* [1]. According to recent analytics of millions of fitness app users, most of them give up their New Year's resolutions to be more physically active only twelve days later [1]. This insight remarkably echoes two prevalent issues: lacking regular physical activity and questionable effectiveness of fitness apps to sustainably motivate users.

The lack of regular physical activity is a pertaining problem for most western societies [2]. Although of high importance for health and well-being, most people are not regularly physically active as recommended [3]. For instance, only 43% of the German population meets the recommended minimum of physical activity in 2018 – a serious downward trend compared to 60% who met the recommendation in 2010 [4]. Initiating and sustaining physical activity is a great challenge for health promotion [2].

Fitness tracking applications and devices aim to address this issue [5, 6, 7] and gain huge public interest [7, 8]. In 2018, about 489 million people already use a fitness

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

tracker [9]. Today, the worldwide market has an estimated revenue of 16 billion USD and is expected to increase in the next years [9]. To motivate people for physical activity, fitness tracking applications ('fitness apps') provide different 'motivational affordances' to the users: self-monitoring, rewards, and social comparison [10, 11].

Although these are expected to benefit individuals by increasing motivation and physical activity, studies draw an overall inconclusive picture about their influence raising controversial discussions about the role these affordances play [5, 6, 7]. It is known that the expected benefits do not unfold to the same extent for every fitness app user: each motivational affordance can be 'motivating' for some – but 'demotivating' for others [5, 6, 7]. Today, however, we lack an understanding as to *why* the motivational affordances are not necessarily of equal benefit for everyone. Understanding the particular effects and causes allows to tailor the motivational affordances and hence to design effective fitness apps that motivate users individually.

Although it is well-known that individuals, and hence fitness app users, can greatly differ in their underlying motives and goals for physical activity [12, 13], little attention has been paid to such motivation-relevant differences of the users and their interplay with motivational affordances in affecting the benefits gained from fitness apps. Therefore, this paper asks: *How do motivational affordances and motivational differences of the users interact in affecting the benefits gained from using fitness apps?*

To provide answers to this question, this paper develops a parsimonious theoretical understanding of how motivational affordances and motivation-relevant user characteristics interact in providing the expected benefits. Drawing upon the key tenets of Achievement Goal Theory [14, 15], we discuss that motivational affordances serve as 'goal structures' that need to be congruent with the 'goal orientations' of the users in order to provide the expected benefits. Quantitative data (N=283) generally lend support for our theoretical considerations, so that this paper contributes to 1) fitness app research by resolving parts of the inconclusive findings about motivational affordances by taking user characteristics into account [5, 6, 7, 16] and 2) by offering a theorizing on motivational affordances [17] explicating their motivation-theoretical characteristics and their interaction with individuals' motivational goals in unfolding their potentials.

The rest of this paper unfolds as follows. Next, we outline the theoretical background of motivational affordances in the context of fitness apps and introduce the key tenets of Achievement Goal Theory. Then we develop our theoretical considerations and detail our hypotheses subsequently. Finally, we lay out our methodological approach and the research results before discussing the findings, implications, and limitations.

2 Theoretical Background

2.1 Motivational Affordances within the Fitness App Context

Fitness apps – such as Strava or Nike+ Running – aim to increase motivation and physical activity. These apps record and document physical activity metrics such as distance, speed, or heart rates and can be used standalone or in combination with devices such as Fitbit wristbands [8]. To augment the recorded data, to induce motivation, and to sustain physical activity, fitness apps provide a set of 'motivational affordances' to the user [5, 6, 7, 16] that build upon self-quantification, gamification, and social network capabilities [16, 18].

The general concept of *'affordances'* is defined within the information systems (IS) research context as "*the possibilities for goal-oriented action afforded to specified user groups by technical objects*" [19, p. 622]. Affordances reflect the potential ways of using IT and aid understanding how the benefits of IT unfold [19]. Using this general notion of affordances in the motivation context, *'motivational affordances'* denote "*the properties of an object that determine whether and how it can support one's motivational needs*" [17, p. 145]. Popular motivational affordances in fitness apps (Table 1) include 'self-monitoring', 'rewards', and 'social comparison' [10, 11].

Table 1. Popular motivational affordances of fitness apps [10]

Affordance	*Definition:* The possibility to…	*Feature examples*
Self-monitoring	… systematically document and observe one's sport behavior	Logs of activity metrics (e.g., time, distance, pulse)
Rewards	… obtain cognitive or virtual rewards for physical activity	Points, badges, trophies
Social comparison	… compare one's performance against others	Leaderboards, competitions, activity reports, profiles

The *self-monitoring affordance* of fitness apps provides the possibility to systematically document and observe one's sports behavior [10] and reflects the cornerstone of the self-quantification [8]. When monitoring themselves, fitness app users seek to observe trends and patterns about their sports behavior. This includes among others whether they are making progress, to ensure that they are maintaining their physical activity or to increase self-awareness about (un)healthy behavior [10]. The *rewards affordance* provides users the possibility to obtain cognitive or virtual rewards for physical activity such as through virtual points or trophies [10] and reflects the gamification aspect of fitness apps [16]. Rewards can be given for achieving self-set activity goals and for making progress (e.g., running a certain distance) but also on normative bases, such as in leaderboards where users obtain trophies for the best sports performance [5]. The *social comparison affordance* reflects the social network aspect [16] that allows fitness app users to compare their performance against other users, for instance through leaderboards, competitions, or others' activity reports and profile pages [10].

However, little is known about the role these motivational affordances play for increased motivation and physical activity. On the one hand, prior research neglected the particular influence of these motivational affordances as fitness apps have been most often examined as a 'whole' [5]. Studies devoted to these particularities, on the other hand, give rise to skepticism about the expected benefits emerging from the motivational affordances. Here, literature reviews draw an overall inconclusive picture reporting positive, neutral, and mixed effects on benefits-related outcomes for the motivational affordances in fitness apps [cf. 5, 6, 7]. As equally accentuated by user stories [e.g., 20], it can be increasingly observed that the achieved benefit of each motivational affordance differs for users. For instance, due to the self-monitoring affordance, users report higher health awareness, higher satisfaction when achieving sports goals, and heightened sports motivation but for others, this affordance is demotivating because of not making visible progress [20, 21]. As for the rewards affordance, studies observed increases in motivation and physical activity, whilst others detected these effects only in short-term or even found no effects at all [5]. The social comparison affordance can be beneficial by inducing fun and affiliation but it also results in peer pressure, negative self-evaluations, or unwanted competition [20, 22].

Whilst research is aware that the expected benefits do not necessarily unfold for all fitness app users to the same extent [5, 6, 7], we do not know a lot about *why* this differs. Scholars point to potential motivation-relevant individual differences of the users [5, 6, 7], such as different motives [16] or goals [18], that can be influential for the motivational affordances used in fitness apps. Thus, taking motivation-relevant user differences into account may provide a more complete understanding of the role motivational affordances play in accounting for the benefits gained from fitness apps. To understand this potential interplay, we draw upon Achievement Goal Theory next.

2.2 Achievement Goal Theory

Human motivation considers the processes that give behavior its energy and direction [12]. Motivation scholars seek to understand the sources of human motivation and their resulting behavior, such as physical activity, suggesting that both the person oneself as well as her/his environment are influential [12, 23]. Achievement Goal Theory (AGT) is a dominant motivation-theoretical framework that considers these dual sources of human motivation and their influence on motivation-relevant cognition, affect, and behavior particularly in the physical activity context [13, 14, 15, 24]. AGT rests on the assumption that individuals are goal-directed beings and achievement goals guide their beliefs, decision making, and subsequent behavior in achievement contexts, such as in the physical activity context [13]. Here, 'achievement' reflects the attainment of a personally or socially valued goal that has meaning for the person, such as improving physical abilities, sports performance, or beating others [13]. AGT hence considers two major achievement goals: *mastery goals* focusing on competence development and *performance goals* focusing on competence demonstration [24]. These two achievement goals are salient within the individual – known as 'goal orientations'– but also within one's social environment – known as 'goal structures' [24].

Goal orientations are dispositional tendencies of the individual [24] where the outlined mastery and performance goals are of centrality. Individuals with a *mastery goal orientation* focus on competence development believing that effort and hard work will lead to competence and mastery [15]. These persons assign value to progress in a self-referenced manner, focusing on learning rather than on the outcome [14, 25]. Thus, fitness app users holding a mastery goal orientation may be eager to improve their physical abilities, to run faster or longer than before, and to observe the progress they made. Individuals holding a *performance goal orientation*, in contrast, define achievement on normative bases and focus on competence demonstration, to show superior ability, to outperform others, and to gain favorable judgments [15]. Fitness apps may be thus motivating and beneficial for users with a performance goal orientation as the social network functionalities allow them to compare themselves with other users, engage in competitions, and to earn trophies for their sports performance.

Goal structures, also referred to as the 'motivational climate' [14], are the achievement-relevant and goal-related emphases within the social environment of the individual [14]. They are caused by environmental practices such as specific messages sent by social actors like teachers, sports coaches, or peers [14, 24]. Equally, goal structures can have an emphasis on mastery and performance goals [24]. A *mastery goal structure* emphasizes improvement and understanding where individuals perceive that effort and learning are valued. A *performance goal structure* stresses relative ability, social comparison, and interpersonal competition [14, 24]. We theorize the role of goal structures in our context subsequently (chapter 2.3).

As goal orientations and goal structures are influential for motivation-relevant outcomes, including heightened motivation and physical activity, linkages between these two constructs have been established [13, 24]. One specific linkage considers the interaction between goal orientations and goal structures in beneficial but also detrimental ways [24]. This interactionist approach proposes a conventional *'matching hypothesis'* suggesting that the most positive outcomes are expected when goal orientations and goal structures are congruent concerning their emphasis on mastery and performance goals [24]: individuals holding a mastery goal orientation are expected to achieve higher motivational outcomes when acting within a mastery goal structure whereas individuals high in a performance goal orientation are best situated within a performance goal structure [24]. Hence, goal orientations and goal structures can reinforce or diminish each other in affecting motivation-relevant outcomes depending upon whether the same goals are emphasized [24].

Next, we discuss how the key tenets of AGT with respect to goal orientations and goal structures with differing emphases on mastery and performance goals as well as their interaction translate into the fitness app context to understand how the benefits of fitness apps to induce motivation and heightened physical activity unfold.

2.3 Theoretical Considerations

Based on these considerations, we aim to use the key tenets of AGT to better understand the role of motivational affordances in promoting the expected benefits of fitness apps.

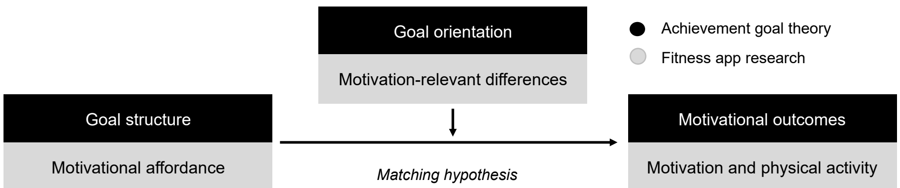


Figure 1. Theoretical considerations

As depicted in Figure 1, we observed parallel lines between fitness app research and AGT. Both fitness app research and AGT are concerned with motivation-relevant outcomes, particularly heightened motivation and increased physical activity. Moreover, fitness app research indicates the necessity to take motivational differences of the users into account which parallels individuals' goal orientations in AGT. Lastly, AGT considers goal structures as an important environmental factor which shares commonalities with the motivational affordances of fitness apps as follows. Information technology, such as a fitness app, is part of an individual's environment including its associated features and capabilities [23]. These features and capabilities, in turn, have certain qualities making them significant for human motivation [23] as posited by the concept of motivational affordances [17]. Motivational affordances can be thus generally understood as environmental factors which are important sources of human motivation [12, 23]. In terms of AGT, motivational affordances act as goal structures with different emphasis on mastery and performance goals.

In short, the *self-monitoring affordance* provides a mastery goal structure given its emphasis on understanding and improvement in a self-referenced manner [10]. The *rewards affordance* can serve as both mastery and/or performance goal structure because sports performance can be either rewarded due to own achievements and improvements but also on normative bases, such as by leaderboards [5, 10]. The *social comparison affordance* serves as a performance goal structure given its emphasis on normative evaluations of one's sports abilities and achievements [10, 16].

In line with the 'matching hypothesis', the motivational affordances are thus expected to interact with users' goal orientations where higher benefits of fitness apps emerge out of a fit between mastery and performance goal emphasis. We detail our arguments subsequently in our hypotheses development.

3 Hypotheses

Because of the inconclusive results about the role motivational affordances play in promoting motivation and physical activity in fitness apps, it is important to understand how and why these benefits do not unfold to the same extent for each user.

In IS research, the benefits gained from IT use are usually labeled as 'net benefits' [26]. As fitness apps aim to enhance motivation and physical activity, net benefits are defined in that context as the extent to which the fitness app has positive impacts on motivation and resulting physical activity. Based on our theoretical considerations just developed (cf. chapter 2.3), we now detail our hypotheses how motivational affordances pronounce mastery and performance goals and thus interact with users' goal orientations in accounting for variations of these net benefits gained.

The *self-monitoring affordance* provides a mastery goal structure to the user [10]. Because it grants the possibility to document sports behavior and to monitor progress in physical activity [8, 10], the self-monitoring affordance focuses on the user's competence and ability and hence emphasizes a mastery goal [14, 24]. As such, the self-monitoring affordance should be particularly beneficial for users holding a mastery goal orientation as it complements their striving to improve their physical abilities and to observe their progress in a self-referenced manner [14, 27]. Recent research in the strand of fitness apps has shown that users holding this mastery goal orientation [18] or pursuing related physical improvement and achievement motives [16] are more inclined towards features that allow evaluation of performance and progress as resembled in the self-monitoring affordance [16, 18]. This evidence suggests that mastery goal orientations match with the self-monitoring affordance making its positive influence even stronger. Fitness app users holding a performance goal orientation, on the other hand, seek to evaluate their sports ability and performance against other users [15]. Here, the self-monitoring affordance does not provide a complementary goal structure that satisfies their performance goals.

H1(a-c): The self-monitoring affordance a) poses a positive relationship with net benefits and this relationship is b) stronger for users high in a mastery goal orientation and c) weaker for users high in a performance goal orientation.

The *rewards affordance* provides a mastery and/or performance goal structure to the user. Rewards can be granted on self-referenced attainments when a user achieved her/his self-set goals such as running 5km, but also on normative bases where a user receives a trophy when her/his performance is better than those of others [5, 10, 16]. Hence, the rewards affordance can be particularly beneficial for users holding a mastery or performance goal orientation. For users with a mastery goal orientation, this affordance values their improvement efforts. Fitness app research tends to support this consideration as users are more attracted by reward-related features of fitness apps when they seek to compare current against past physical conditions or aim at reaching a particular activity goal [18]. For users with a performance goal orientation, this affordance makes their achievements visible to others and awards outperforming other users [5, 10, 16]. Here, fitness app research equally lends support as users high in a performance goal orientation place higher importance on reward-related features [18]. Thus, this affordance supplements mastery and/or performance goal orientations.

H2(a-c): The rewards affordance a) poses a positive relationship with net benefits and this relationship is stronger for users b) high in a mastery goal orientation and c) high in a performance goal orientation.

The *social comparison affordance* provides a performance goal structure to the users as it promotes interpersonal competition and public evaluation [14, 28] through features that allow fitness app users to compare their performance against others users' performances [10]. This affordance is thus expected to be most beneficial for users holding a performance goal orientation [24]. Users holding a performance goal orientation describe their achievements on normative bases and seek to become better than others [15]. The social network capabilities of fitness apps allow to observe other users' profile pages and activities or to enter competitions where performances are displayed within leaderboards [10]. The social comparison affordance should thus reinforce their striving for interpersonal performance comparisons. Research indicates that fitness app users pursuing social motives for engaging in sports, such as affiliation and recognition from others, as well as users pursuing competitive motives are more likely to draw upon features allowing performance comparisons against others as well as on features where they can receive recognition from their social network, such as 'likes' [16]. Similarly, performance-oriented users assign higher importance to social network features including other users' activity logs and leaderboards [18]. Mastery-oriented users, in contrast, should receive fewer benefits from the social comparison affordance because those users focus on self-referenced achievements.

H3(a-c): The social comparison affordance a) poses a positive relationship with net benefits and this relationship is b) weaker for users high in a mastery goal orientation and c) stronger for users high in a performance goal orientation.

4 Research Method

To test our hypotheses, empirical data with specific characteristics is needed. First, as the functionalities vary between apps [29], a single fitness app providing our focal affordances is required to prevent outside effects. Second, to assess the effects of the 'social comparison' affordance, users need to be connected with other users. Third, the fitness app needs to target the same class of activities, such as cardio-intense sports like running or cycling, to assess the achieved benefits (motivation and physical activity).

We chose the fitness app 'Strava' (www.strava.com), which meets above-mentioned criteria and anecdotal user stories indicate varying benefits [e.g., 20]. As a large number of users is needed, mainly to test interaction effects, we recruited respondents using an online panel of Amazon Mechanical Turk that gained widespread attraction as a viable and reliable source for research [30]. Following recent guidelines [30], we restricted our sample to the United States as such responses provide reliable results similar to regular consumer panels [30, 31]. Our study was conducted in May 2018.

As depicted in Table 2, our measurement instrument was derived from prior literature. Items for motivational affordances [10] were adapted to our context. Goal orientation items were assessed by instructing respondents to think about their sports and exercises [25, 27]. Items for net benefits were based on our definition, compiled from related studies [32], and adapted. We used a 7-point Likert scale for all items.

Table 2. Measurement instrument

Construct	*Items*
Affordances	*When I use Strava, I use features that allow me…*
Self-monitoring [10]	… to monitor my sport behavior. … to keep track of my exercise activities. … to record my physical activities.
Rewards [10]	… to make my physical activity rewarded. … to get more rewards if I try harder. … to earn virtual rewards as a token for my efforts in physical activity.
Social comparison [10]	… to compare my performance with the performance of others. … to compare myself with others regarding what I have accomplished in exercising. … to find out how I am doing in exercise compared to what others have done.
Mastery goal [25, 27]	It is important to me to perform as well as I possibly can. I prefer challenging goals so that I'll improve a great deal. I am willing to take on a difficult challenge if it helps me reach my goals.

Performance goal [25, 27]	It is important for me to perform better than others. It is important to me to do well compared to others. To be honest, I really like to prove my abilities to others.
Net benefits [32]	Using Strava has helped me being physically active. I gained motivation to exercise from using Strava. I exercise more since using Strava. Using Strava makes my sport more enjoyable.

After data collection and preparation, 283 responses were eligible for data analysis [30].[1] The dataset is characterized as follows: 60.4% are male and the average age is 32.4 years (SD 7.96 years). On average, participants use Strava already for 11.5 months (SD 14.1 months) and when it comes to sports, 63.3% use Strava 'often' or 'always'. Participants' number of followers in Strava is at a median of 16 followers. Asking about their general frequency of performing sports, 78.4% answered with 'several times a week' or 'almost every day'. As such, our sample consists of quite active sports people resonating with recent literature and studies [16, 18, 29].

5 Data Analysis

The data was subsequently transferred into Structural Equation Modelling using Partial Least Squares [33] with the software SmartPLS 3. We analyzed the measurement model including common method bias before evaluating the hypothesized relationships [33].

All constructs were modeled using reflective measurements. Evaluation involves indicator and construct reliability as well as discriminant validity (Table 3) [33].

Table 3. Measurement model evaluation

			Discriminant validity					
Construct	*CR*	*AVE*	1	2	3	4	5	6
1. Benefits	0.904	0.703	0.838					
2. Monitor.	0.863	0.678	0.667	0.824				
3. Rewards	0.903	0.756	0.397	0.237	0.870			
4. Soc. comp.	0.900	0.750	0.434	0.277	0.595	0.866		
5. Mast. goal	0.862	0.675	0.616	0.537	0.359	0.470	0.822	
6. Perf. goal	0.907	0.765	0.293	0.133	0.490	0.581	0.452	0.875

Indicator reliability was achieved as all item loadings are greater than 0.707 (0.781–0.892) and are significant ($p<0.001$). *Construct reliability* was achieved as values for Composite Reliability (CR) and for Average Variance Extracted (AVE) are above 0.7

[1] We received 624 responses whereby 514 passed the screening criteria of currently using Strava. We then removed all non-unique responses based on IP addresses and Worker IDs (N=425). To evaluate the 'social comparison' affordance, only participants who indicated being connected with other users were retained (N=293). Finally, responses with failed attention checks and more than five missing answers were dropped (N=283).

and 0.5. Construct reliability is further supported by values for Cronbach's Alpha between 0.760 and 0.859. *Discriminant validity* is supported as 1) each item loads highest on its designated construct, 2) as the Fornell-Larcker criterion is fulfilled given that the inter-variable correlations are smaller than the root of the corresponding AVE as demonstrated in the diagonal lines [33], and 3) as the heterotrait-monotrait ratio with the highest value of 0.81 is below 0.85 [34].

Common method bias (CMB) can be of concern when using self-reported data obtained through a single method [35]. To mitigate the potential influence, we stressed anonymity, the academic purpose, that there are no wrong or right answers, and randomized item ordering [35]. To evaluate its presence in our data, two tests were conducted. Results of Harman's single factor test indicate that 39.15% of the variance is attributed to one single factor which is not the majority [35]. Performing an additional test [36] by entering a CMB factor into the model containing all items and observing its respective influence on each construct, the resulting ratio of 1:685 is much smaller compared to prior research [36]. Thus, CMB is not a concern in our data.

For hypotheses testing, we first analyzed the motivational affordances on net benefits without the interactions (Figure 2). We obtained an R^2 of 52.7% for net benefits as significantly predicted by the three motivational affordances with medium to large effect sizes [33]. The results hence support hypotheses H1a, H2a, and H3a.

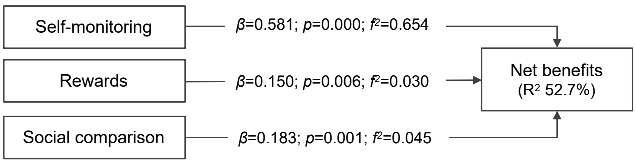


Figure 2. Direct effects of motivational affordances on net benefits

Next, we focused on the interactions of each motivational affordance with mastery and performance goal orientations in isolation. As interaction effect sizes are usually small, a conservative interpretation is 0.005 ('small'), 0.01 ('medium'), and 0.025 ('large') [37]. To aid interpretation, we plotted the significant interactions (Figure 3).

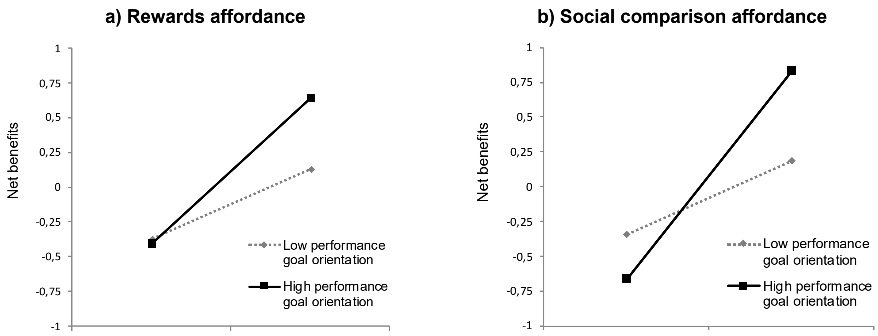


Figure 3. Interactions of a) rewards and b) social comparison affordance with performance goals

For *self-monitoring*, we did not observe significant interactions with mastery goal orientations (β=-0.070; p=0.091; H1b not supported) nor with performance goal orientations (β=-0.034; p=0.590; H1c not supported) on net benefits. For *rewards*, we found a significant positive interaction with performance goal orientations with a large effect size (β=0.138; p=0.013; f^2=0.035; H2c supported) but not with mastery goal orientations (β=0.010; p=0.879; H2b not supported) on net benefits. Lastly, for *social comparison***,** we found a significant positive interaction with performance goal orientations with a large effect (β=0.243; p=0.000; f^2=0.135; H3c supported) but not with mastery goals (β=0.048; p=0.344; H3b not supported) on net benefits. For all models analyzed, we obtained good model fits with SRMR values of 0.062 to 0.073, which are below the threshold of 0.08. We discuss the results next.

6 Discussion

Our research was motivated by the lack of regular physical activity prevailing in most western societies [2] and the inconclusive results about fitness apps and their motivational affordances to increase motivation and physical activity [5, 6, 7]. To resolve the inconsistent findings of prior research, we paid closer attention to the interplay of the self-monitoring, rewards, and social comparison affordances and users' motivation-relevant goal orientations. A quantitative study with 283 users of the fitness app 'Strava' provided support for five of our nine hypotheses. We discuss our findings next before laying out the research contributions, implications, and limitations.

The empirical data highlights the influential role of the three motivational affordances that jointly accounted for 52.7% of the variations of the benefits gained. Self-monitoring emerged to be most influential, which is not surprising as this affordance allows physical activity documentation that is integral to fitness app use [10, 29]. Rewards and social comparison appeared to be less influential, which further increased our interest to take users' goal orientations into consideration. We argued that the influence of motivational affordances varies in dependence upon users' mastery and performance goal orientations. Albeit we only found two significant interactions, these provide meaningful insights for the rewards and social comparison affordances. Because these affordances emphasize a performance goal structure, mostly through leaderboards, they complement particularly performance-oriented users' natural striving for becoming better than others [15] by making their efforts comparable and rewarded on normative bases. Although mastery goal orientations were expected to be a moderating factor as well, we did not obtain the empirical support. This absence can be explained in two ways. First, mastery and performance goal orientations are not mutually exclusive so that individuals can pursue both goals at the same time [13]. Second, Strava positions itself as 'the social network athletes' and our sample consists of quite 'athletic' users holding an above-average mastery goal orientation with little variation (mean 5.52, SD 0.93). As a result, the self-monitoring affordance poses such a strong direct effect that there is little room for variations stemming from users' goal orientations. The rewards affordance in Strava is built on the social network capabilities, particularly leaderboards, so that rewards are here much more based on normative grounds stressing performance goals rather than mastery goals.

Based on the empirical findings discussed so far, we generally note solid support for our theorized interplay of motivational affordances and users' motivation-relevant goals so that our research makes the following contributions to literature and theory.

First, our paper contributes to *fitness app research* [e.g., 5, 6, 7, 16]. As outlined above, little is known about the role motivational affordances of fitness apps play in promoting motivation and physical activity. Parts of this inhibited understanding are caused because prior research mostly examined fitness apps as 'whole' without paying the necessary attention to the particular roles the motivational affordansces play [5]. Thus, we first revealed that the reported benefits of heightened motivation and increased physical activity are generally determined by self-monitoring, rewards, and social comparison affordances. Moreover, prior research also raised awareness that each motivational affordance is not necessarily beneficial for every user – but little has been put forth to understand *why* it can differ [5, 6, 7]. To provide the needed explanations, we took motivation-relevant differences of the users, namely goal orientations, into consideration. We found that the benefits gained from the rewards and social comparison strongly depend upon the goal orientation pursued. Thus, by considering these motivation-relevant differences of the users, our paper resolves parts of the inconclusive findings of prior research [cf. 5, 6, 7]. Our considerations developed here enable further research to better understand the motivational affordances of fitness apps, such as their role in causing psychological need satisfaction and frustration [22] or in promoting continued fitness app use [38].

Second, our paper contributes by *theorizing about motivational affordances and individual differences* [17]. Although motivated by the inconsistent findings in our particular fitness app context, we developed a rather general theoretical account of motivational affordances. As detailed in chapter 2.3 and 3, we noted parallel lines with the key tenets of AGT concerning a) the consequences of interest, b) individual differences in terms of goal orientations, and c) the role of motivational affordances in terms of providing goal structures. This approach provides a refined understanding of the motivation-relevant characteristics of motivational affordances: they act as goal structures with differing emphases on mastery and performance goals. Because of this characterization, our theorizing provides explanations as to why motivational affordances interact with individual goal orientations resulting in variations of the motivational consequences. Our theorizing hereto considers that the motivational affordances and users' goal orientations need to be congruent, sharing the same emphasis on mastery or performance goals. When congruent, the positive consequences amplify yet deteriorate when incongruent. As such, our theorizing provides research with a better understanding of the interplay between motivational affordances and individual, motivation-relevant characteristics. Without this cognizance, researchers may otherwise draw insufficient or misleading conclusions about the role motivational affordances play in achieving the desired outcomes. As such, our theorizing is also useful for other contexts where motivational affordances are expected to favor anticipated outcomes, such as for organizational collaboration systems where similar affordances aim to increase employees' knowledge contributions [39].

For practice, our paper guides the application of motivational affordances in fitness apps. Fitness app vendors need to be aware that users' motivational goal orientations can greatly differ and that motivational affordances are thus not necessarily universally effective in promoting motivation and physical activity. The exposure to motivational affordances, particularly to rewards and social comparison, needs to be tailored to the goal orientations of the users. To achieve this, users' motivational goals should be assessed during initial app interactions in order to expose those affordances that provide the best fit. Then, users are expected to achieve higher benefits for motivation and physical activity so that the expected potentials of fitness apps unfold.

Despite these contributions and implications, we acknowledge some limitations. We focused on one single app offering the focal motivational affordances targeting rather intense cardio sports such as running or cycling. Although fitness app users in this context are already exercising [29], the goals and motives users pursue still differ accounting for differences in the use of cardio-targeting fitness apps [16, 18] and resulting benefits. Likewise, issues concerning the effects of motivational affordances, particularly rewards and social comparison, have also been reported in other fitness app contexts in which users are generally less motivated and in which apps target less intense physical activities such as walking [22]. Thus, to provide further confidence in the theoretical considerations developed here, future research should be carried out across a variety of fitness apps and different target groups, particularly to examine the influence of mastery goals. In this vein, our inquiry is bound to fitness apps offering social network capabilities. Although offered by many popular fitness apps, the app landscape is highly diverse and there are also fitness apps without a 'social dimension' [29], that need to be studied and compared. Hereto, other affordances such as 'exercise guidance' are available depending upon the fitness apps of interest [10, 29] and which can provide further insights. Equally, future research should also investigate other sorts of benefits achieved, such as improved health awareness, to better understand the effects of the motivational affordances provided.

References

1. Whipple, T.: Quitters' day puts new year's resolutions to the test, https://www.thetimes.co.uk/article/quitters-day-puts-new-year-s-resolutions-to-the-test-29mr0vn3g (Accessed: 2018-07-09)
2. WHO: Global recommendations on physical activity for health, http://www.who.int/dietphysicalactivity/publications/9789241599979/en/ (Accessed: 2018-08-14)
3. Penedo, F.J., Dahn, J.R.: Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Current Opinion in Psychiatry 18, 189-193 (2005)
4. Froböse, I., Biallas, B., Wallmann-Sperlich, B.: DKV-Report 2018: Wie gesund lebt Deutschland?, https://www.ergo.com/-/media/ergocom/pdf-mediathek/studien/dkv-report-2018/dkv-report-2018.pdf (Accessed: 2018-08-09)
5. Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., Hides, L.: Gamification for health and wellbeing: A systematic review of the literature. Internet Interventions 6, 89-106 (2016)

6. Hamari, J., Koivisto, J., Pakkanen, T.: Do Persuasive Technologies Persuade? - A Review of Empirical Studies. In: International Conference on Persuasive Technology, pp. 118-136. (2014)
7. Orji, R., Moffatt, K.: Persuasive technology for health and wellness: State-of-the-art and emerging trends. Health Informatics Journal 24, 66-91 (2018)
8. Lupton, D.: The Quantified Self. Polity Press, Malden, MA (2016)
9. Statista: eServices Report 2018 – Fitness, https://de.statista.com/statistik/studie/id/36673/dokument/fitness-outlook/ (Accessed: 2018-07-23)
10. Rockmann, R., Gewald, H.: Activity Tracking Affordances: Identification and Instrument Development. In: Pacific Asia Conference on Information Systems. Yokohama, Japan (2018)
11. Lister, C., West, J.H., Cannon, B., Sax, T., Brodegard, D.: Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2, e9 (2014)
12. Reeve, J.: Understanding Motivation and Emotion. John Wiley & Sons, Inc., New York (2005)
13. Roberts, G.C., Treasure, D.C., Conroy, D.E.: Understanding the Dynamics of Motivation in Sport and Physical Activity: An Achievement Goal Interpretation. In: Tenenbaum, G., Eklund, R.C. (eds.) Handbook of Sport Psychology, pp. 3-30. John Wiley & Sons, Inc., Hoboken, New Jersey (2007)
14. Ames, C.: Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology 84, 261-271 (1992)
15. Maehr, M.L., Zusho, A.: Achievement Goal Theory: The Past, Present, and Future. In: Wentzel, K.R., Wigfield, A. (eds.) Handbook of Motivation at School, pp. 77-104. Routledge, New York, NY (2009)
16. Stragier, J., Vanden Abeele, M., De Marez, L.: Recreational athletes' running motivations as predictors of their use of online fitness community features. Behaviour & Information Technology 37, 815-827 (2018)
17. Zhang, P.: Motivational Affordances: Reasons for ICT Design and Use. Communications of the ACM 51, 145-147 (2008)
18. Hamari, J., Hassan, L., Dias, A.: Gamification, quantified-self or social networking? Matching users' goals with motivational technology. User Modeling and User-Adapted Interaction 28, 35-74 (2018)
19. Markus, M.L., Silver, M.S.: A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole's Concepts of Structural Features and Spirit. Journal of the Association for Information Systems 9, 609-632 (2008)
20. Barratt, P.: Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling. Health & Place 46, 328-336 (2017)
21. Lee, V.R., Drake, J.: Digital Physical Activity Data Collection and Use by Endurance Runners and Distance Cyclists. Technology, Knowledge and Learning 18, 39-63 (2013)
22. Kerner, C., Goodyear, V.A.: The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on Fitbits with adolescents. American Journal of Health Education 48, 287-297 (2017)
23. Zhang, P.: Toward a Positive Design Theory: Principles for Designing Motivating Information and Communication Technology. In: Avital, M., Bolland, R., Cooperrider, D. (eds.) Designing Information and Organizations with a Positive Lens, pp. 45-74. Emerald Group Publishing Limited (2007)

24. Murayama, K., Elliot, A.J.: The joint influence of personal achievement goals and classroom goal structures on achievement-relevant outcomes. Journal of Educational Psychology 101, 432 (2009)
25. Elliot, A.J., McGregor, H.A.: A 2×2 Achievement Goal Framework. Journal of Personality and Social Psychology 80, 501-519 (2001)
26. Zhang, X., Venkatesh, V.: From Design Principles to Impacts: A Theoretical Framework and Research Agenda. AIS Transactions on Human-Computer Interaction 10, 105-128 (2018)
27. Conroy, D.E., Elliot, A.J., Hofer, S.M.: A 2x2 Achievement Goals Questionnaire for Sport: Evidence for Factorial Invariance, Temporal Stability, and External Validity. Journal of Sport and Exercise Psychology 25, 456-476 (2003)
28. Morris, R.L., Kavussanu, M.: Antecedents of approach-avoidance goals in sport. Journal of Sports Sciences 26, 465-476 (2008)
29. Higgins, J.P.: Smartphone Applications for Patients' Health and Fitness. The American Journal of Medicine 129, 11-19 (2016)
30. Jia, R., Steelman, Z.R., Reich, B.H.: Using Mechanical Turk Data in IS Research: Risks, Rewards, and Recommendations. Communications of the Association for Information Systems 41, 301-318 (2017)
31. Steelman, Z.R., Hammer, B.I., Limayem, M.: Data Collection in the Digital Age: Innovative Alternatives to Student Samples. MIS Quarterly 38, 355-378 (2014)
32. Pfeiffer, J., von Entress-Fuersteneck, M., Urbach, N., Buchwald, A.: Quantify-Me: Consumer Acceptance of Wearable Self-Tracking Devices. In: European Conference on Information Systems. Istanbul (2016)
33. Chin, W.W.: The Partial Least Squares Approach to Structural Equation Modeling. In: Marcoulides, G.A. (ed.) Modern Methods for Business Research, pp. 295-336. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (1998)
34. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science 43, 115-135 (2015)
35. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. Journal of Applied Psychology 88, 879–903 (2003)
36. Liang, H., Saraf, N., Hu, Q., Xue, Y.: Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. MIS Quarterly 31, 59-87 (2007)
37. Kenny, D.A.: Moderation, http://davidakenny.net/cm/moderation.htm (Accessed: 2018-03-20)
38. Suh, A.: Sustaining the Use of Quantified-Self Technology: A Theoretical Extension and Empirical Test. Asia Pacific Journal of Information Systems 28, 114-132 (2018)
39. Suh, A., Wagner, C.: How gamification of an enterprise collaboration system increases knowledge contribution: an affordance approach. Journal of Knowledge Management 21, 416-431 (2017)

Gamification in Health Behavior Change Support Systems - A Synthesis of Unintended Side Effects

Manuel Schmidt-Kraepelin[1], Scott Thiebes[1], Stefan Stepanovic[2], Tobias Mettler[2], and Ali Sunyaev[1]

[1] Karlsruhe Institute of Technology,
Department of Economics and Management, Karlsruhe, Germany
{manuel.schmidt-kraepelin,scott.thiebes,sunyaev}@kit.edu
[2] University of Lausanne, Lausanne, Switzerland
{stefan.stepanovic,tobias.mettler}@unil.ch

Abstract. Gamification has become a popular and promising tool to positively impact the usage of health behavior change support systems (HBCSSs). Fun and engaging components are purposefully integrated in the design of HBCSSs in an effort to encourage users to employ the system in a more regular manner or over a longer period of time. Although extant research has made extensive efforts to understand the psychological and behavioral outcomes of gamification, its potential unintended side effects have been mostly neglected. We approached this gap by reviewing 33 articles on gamification in HBCSSs. We identified 16 potential unintended side effects in five categories. By taking a critical view, our research contributes to a more nuanced approach to gamification, which helps to understand how it can be utilized as a valuable tool for developers that motivates and does not harm users of HBCSSs.

Keywords: Gamification, Health Behavior Change, Literature Review, Side Effects, Negative Effects

1 Introduction

In today's Western societies the main health challenges and most important risks for mortality have been shifted from pre-modern health risks such as malnutrition and poor water quality to health risks that are generated by the modern world itself, like high blood pressure, tobacco use, high blood glucose levels, obesity, and high cholesterol levels [1]. Thus, major public health concerns are often directly linked to people's individual health behavior and modern lifestyle habits such as sedentary living, chronic stress, and intake of energy-dense foods [1]. Trying to tackle the challenge of unhealthy modern lifestyles, health behavior change support systems (HBCSSs) are a promising approach to positively influence people's health behavior [2-4]. Typical examples of such HBCSSs include systems that aim to foster healthy eating habits [5, 6], systems that motivate their users to be more physically active [7, 8], or systems that help to properly manage chronic diseases [9, 10]. HBCSSs come in many forms, such as wearables, mobile apps or dedicated software. However, recent research suggests that

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

people often do not use HBCSS frequently and over a sustained period of time [11]. As a consequence, the desired beneficial health behavior may not be maintained. Trying to address this issue, developers of HBCSS frequently employ gamification. Gamification refers to the implementation of game elements in non-game contexts [12].

In HBCSSs gamification is primarily applied to make the usage of a HBCSS more engaging and fun, thus encouraging users to use the HBCSS more regularly or promoting the completion of certain health-related activities [2, 4]. However, designing meaningful and effective gamification is associated with high complexity and extensive resource requirements [13]. In fact, researchers have repeatedly criticized practitioners and fellow researchers for oversimplifying gamification by superficially implementing single game elements such as points or badges [14]. Adding to that, recent research suggests that gamification approaches that do not consider the overall application context and gamification concept are less effective than more holistic approaches to gamification [15]. Mindless approaches to gamification might not only lack effectiveness in increasing users' motivation and engagement but additionally cause unintended side effects that could counteract the positive effects of gamified systems or even harm their users [16, 17]. Especially in a context like HBCSSs, where systems are intended to have a positive impact on peoples' health behavior, unintended side effects of gamification can have serious negative influences on users' health outcomes (e.g., by unintentionally incentivizing wrong exercising). Thus, developers of gamified HBCSSs need to be aware of potential unintended side effects in order to incorporate suitable strategies to address them into the design process. Within this research, we hence aim to answer the following research question:

RQ: What unintended side effects may occur when implementing gamification in HBCSSs?

Past research on gamification has focused on investigating the positive effects of gamification on psychological and behavioral outcomes [18, 19] or proposing theoretically grounded frameworks for designing specific gamified systems [e.g., 20]. Risks and negative aspects of gamification were only treated as side notes [e.g., 1, 21]. Some researchers have started to investigate negative aspects of gamification in general [16] or within educational systems [17]. However, those studies are based on the analysis of secondary literature and do not consider the special context of HBCSSs. This is rather problematic since (1) unintended side effects of (gamified) HBCSSs yield potential to seriously harm users' health and thus substantially differ from side effects in less serious contexts and (2) the investigated secondary literature did not thoroughly elaborate on negative aspects of gamification. In consequence, research still lacks a comprehensive overview of unintended side effects of gamification in HBCSSs. Although existing studies make first valuable contributions to the research field, a synthesis of literature is necessary to understand unintended side effects of gamification in particular with regard to specifics of HBCSSs and the serious contexts of HBCSSs.

To answer our research question, we conduct a structured review of literature to identify and analyze relevant academic publications. In particular, we review those publications that discuss and elaborate on potential unintended side effects of applying gamification to HBCSSs. An overview and explanation of unintended side effects of gamification in HBCSSs helps to (1) guide developers of gamified HBCSSs in identifying potential risks within their gamification concepts and (2) raise awareness that gamification is not a silver bullet which creates positive outcomes all by itself and without extensive design considerations.

This paper proceeds as follows. The next section provides an outline of gamification in HBCSSs as well as an overview of research on unintended side effects of gamification. Section three describes our research approach, while section four presents our results. We discuss our results in section five and briefly conclude our paper in section six.

2 Background

2.1 Gamification in Health Behavior Change Support Systems

Literature provides two prevailing definitions for gamification. Huotari and Hamari [22] refer to gamification as the process of enhancing services with motivational affordances for gameful experiences. Hamari et al. [19] advanced this conceptualization by introducing the three essential concepts in gamification research (i.e., implemented motivational affordances, resulting psychological outcomes, and further behavioral outcomes). Deterding et al., define gamification as "the use of game design elements in non-game contexts" [12]. Popular game elements used in gamification include points, badges, leaderboards, and time constraints [12]. In general, gamification aims to utilize peoples' growing passion for games to positively influence their personal motivation or perception concerning a selected action in order to make it more engaging and fun [2, 21, 23]. However, it is important to differentiate gamification and serious games. Gamified systems are no full-fledged games. In fact, game elements in gamified systems are only means to foster certain behaviors and not the main object of the system [24]. Serious games, on the other hand, are fully-developed games that serve specific non-entertainment purposes [24].

Extant research has made extensive efforts to investigate psychological and behavioral effects of gamification in various research fields such as education [25, 26], crowdsourcing [27], or enterprise systems [21]. Among these research fields, HBCSSs have emerged as one of the most relevant application areas for gamification [1, 28]. In HBCSSs, gamification is primarily applied for motivating individuals to continue using the systems more regularly or promoting the completion of activities or tasks that are associated with positive health outcomes [2, 4]. According to extant literature, there are three major groups of use contexts for gamification in HBCSSs [2]: (1) Individual lifestyle habits (e.g., fitness, food consumption, unhealthy habits), (2) chronic disease management (e.g., diabetes, cancer), and (3) support of health professionals (e.g., for educational purposes or daily habits). However, as we focus our research on unintended side effects concerning patients and users aiming to improve their health status as well as the fact that a majority of HBCSSs in group three can also be classified as educational systems, we concentrate our analysis on such HBCSSs belonging to group one and two.

2.2 Unintended Side Effects of Gamification

Extant research on gamification has primarily focused on investigating intended psychological and behavioral effects on gamification. Two studies exist that focus on shedding light on potential unintended side effects of gamification. Hyrynsalmi and Kimppa [16] conducted a meta study and reviewed existing literature reviews on gamification concerning negative impacts. They classify their results in two main groups: (1) Limiting issues and (2) harmful issues. However, they also state that most reviewed studies had only little if any discussion on the negative effects of gamification. As a result, their data basis is scarce and lacks an in-depth discussion of side effects and potential consequences. Second, Toda et al. [17] reviewed literature on gamification in education and identified four negative effects (i.e., indifference, loss of performance, undesired behavior, and declining effects). However, the results of their study are only applicable to the context of educational systems. In addition to these two studies, some research has dealt with risks or negative effects of gamification as a side note. For example, Thiebes et al. [21] have discussed four risks of gamification in information systems, and Johnson et al. [1] outline some negative aspects of gamification found in studies in health and well-being. Furthermore, Kim and Werbach [29] elaborate on ethical issues in applying gamifications such as potential for manipulation and exploitation. In summary, to the best of our knowledge, no study exists that focuses on identifying and purposefully reasoning on unintended side effects of gamification in HBCSSs.

3 Structured Literature Review

3.1 Data Collection

For the identification of publications discussing potential unintended side effects of gamification in a HBCSSs context, we applied a systematic online literature database search following the guidelines by Levy and Ellis [30]. We thus searched the scientific databases IEEE Xplore, ProQuest, AIS Electronic Library, ACM Digital Library, EBSCO Host, and ScienceDirect using the following search string: *TITLE-ABSTR-KEY (gamif*) and TITLE-ABSTR-KEY(health* OR medic* OR life* OR fitness OR well-being) and TITLE-ABSTR-KEY(risk* OR danger* OR peril* OR effect* OR negative* OR disadvantage*).*

Where possible, our search was limited to peer-reviewed publications published in 2010 or later, since gamification only gained widespread recognition by researchers and practitioners in 2010. The database search yielded a total of 212 publications, excluding duplicates. Two researchers separately assessed the relevance of each article by utilizing predefined exclude criteria. In this process, we excluded ten articles not written in English, seven articles that were not peer-reviewed, 140 articles that had no focus on gamification in HBCSSs (i.e., they dealt with related concepts such as serious games or researched gamification in a non-healthcare context), and 43 articles that did not discuss any unintended side effect of gamification. In addition, a forward and backward search was conducted on the twelve relevant articles which lead to the identification of three additional relevant publications and a set of 15 relevant articles. In a second step and in an effort to integrate scientific literature from health and medical perspectives, we searched the scientific database PubMed using our research string. Our PubMed search yielded a total of 145 additional unique publications. By screening these publications against our exclusion criteria, we excluded 104 publications that had no focus on gamification in HBCSSs and 23 publications that did not discuss any unintended side effects of gamification. Through searching PubMed, we identified 18 additional relevant publications, which led to a final set of 33 relevant publications.

3.2 Concept-centric Data Analysis

To identify unintended side effects of gamification in health & well-being, we conducted a manual content analysis. Two researchers independently coded the 33 articles with regard to unintended side effects using an open coding approach [31]. As suggested by Strauss and Corbin [31], during open coding, the data was broken down into discrete parts (i.e., text passages), closely examined, compared for similarities and differences, and coded with regard to the phenomena as reflected in the data. The results were iteratively reviewed and discussed with a third researcher to consolidate them. To improve the explanatory power of our results and to find semantically coherent groups, the identified side effects were grouped into categories if they were logically related to the same subject.

4 Results

Our review of relevant literature yielded a total of 16 potential unintended side effects due to the application of gamification in HBCSSs, which we grouped into five categories of unintended side effects. Table 1 provides an overview of the categories of unintended side effects. It also highlights for which unintended side effects we were able to find empirical support within the reviewed literature and which side effects are specific to the HBCSSs context. We describe each unintended side effect in detail below.

4.1 Adverse Motivational Outcomes

Undermining Intrinsic Motivation. Researchers often argue that gamification aims to foster users' intrinsic motivation in order to make using gamified systems more engaging and fun [32]. However, gamified HBCSSs sometimes tend to focus on extrinsic motivation and thereby even corrupt and undermine intrinsic motivation for positive health behavior change [33, 34]. As a result, users' health behavior may become dependent on the presence of the gamified HBCSSs and their motivation may immediately decrease once the extrinsic rewards are not available anymore [1, 33]. Attig and Franke [33], for example, showed in their study that motivation for physical activity can become dependent on the presence of an activity tracker as well as the related game elements and decreases in case the tracker is not available.

Table 1. Overview of unintended side effects.

Class of Unintended Side Effects	*Unintended Side Effect*	*Sources*	*Empirical support*	*HBCSS specific*
Adverse Motivational Outcomes	Undermining Intrinsic Motivation	[13, 33-42]	[33-35, 38, 40, 41]	No
	Motivation Decreasing Over Time	[15, 35, 38, 40, 43-48]	[35, 40, 43, 45-48]	No
	Unfulfilled Expectations	[36, 45, 49, 50]	[45, 49]	No
Informational Noise	Distraction from Health Purpose	[13, 15, 34, 41, 45, 51, 52]	[15, 45, 51, 52]	Yes
	Trivializing the Health Context	[1, 13, 15, 24, 36, 41, 53]	-	Yes
	Reduced Usability	[13, 15, 45, 54]	[45]	No
Reduced Integrity of Exercise	Cheating the Self	[37, 55]	-	Yes
	Rewarding Incorrect Execution	[34, 36, 56]	-	Yes
	Overuse	[13, 55, 57, 58]	[57]	Yes
Demoralization of Users	Cheating Others	[13, 37, 55, 59]	-	No
	Overemphasized Peer Pressure	[37, 38, 41, 57, 60]	[38, 41, 57, 60]	No
	Exaggerated Punishment	[37, 40]	-	No
	Feeling of Manipulation	[36, 42, 49, 57]	[49, 57]	No
	Discouragement Due to Failure	[36, 61]	-	No
Overstepping Boundaries	Privacy Infringements	[13, 36, 37, 40, 44, 55, 60]	[40, 60]	Yes
	Fostering Behavior that Harms Third Parties	[58]	[58]	No

Motivation Decreasing Over Time. An often-discussed problem with gamification is novelty effects. Novelty effects describe a situation in which users are often curious and enthusiastic for gamification at first, as it is visually appealing and something they did not experience before [15]. However, interest and enthusiasm for gamification most likely decrease in the long run when these novelty effects wear off. With it, motivation to perform healthy behaviors could decrease as well, ultimately dropping below the initial level of motivation. El-Hilly et al. [40], for instance, describe this effect in their study of a gamified HBCSS for smoking cessation where participants exhibited monotony and decreased levels of engagement as they perceived achievements as repetitive.

Unfulfilled Expectations. Gamification of HBCSSs might raise high expectations by claiming to bring fun and engagement to health activities while maintaining therapeutic effectiveness of the system. If these expectations are not met, users might be disappointed, which could lead to decreasing levels of satisfaction. To this end, Lumsden et al. [49] report that participants of their study of different web-based cognitive testing systems were disappointed of a task that had the graphical impression of a game but did not offer any actual gameplay.

4.2 Informational Noise

Distraction from Health Purpose. Sardi et al. [15] point out that gamification concepts sometimes tend to not provide "a tangible health-driven meaning in terms of the user's competence and health skills" and that game mechanics are "sometimes wrongly located on the application's display". As a result, it can be difficult for users to identify a link between the gamification concept and their health behavior and they do not understand the purpose of certain game mechanics. This effect can lead to a distraction from the core health behavior elements of the system and thus reduce the overall system's efficacy [34]. For example, Boendermaker et al. [45] conclude in their study that their gamified HBCSSs for attentional bias modification in the context of alcohol consumption contained distracting game elements that negatively influenced the systems efficacy.

Trivializing the Health Context. Developers of gamification concepts frequently aim to design visual appearances that resemble existing games [21]. Thus, gamification design is often colorful and eye-catching [1]. In some cases, exaggerated visual design might lead to perceptions that important health topics, which deserve a serious and professional tone, might be trivialized and that gamification is more of a marketing gimmick than a serious tool that supports health behavior change [36]. As a result, recent studies reported that some health professionals shy away from participating in designing gamified HBCSSs as they worry about their credibility and respect among patients [15].

Reduced Usability. Introducing gamification to a HBCSS is always associated with adding new possibilities of system interaction for the user. As a result, established paths of human-system interaction may change and users' might initially be confused as the system does not longer work the way it used to. Gamifying HBCSSs adds an additional layer of interaction complexity and, thus, might lead to an initial decrease in usability [15]. For example, Boendermaker et al. [45] observed that participants in a gamified intervention differed with regard to speed and accuracy of responses in comparison to a non-gamified intervention. They ascribe this observation to the more complex nature of the gamified intervention.

4.3 Reduced Integrity of Exercise

Cheating the Self. Gamifying HBCSSs can open the door to cheating and exploiting. Especially users that have higher interest in achieving game rewards than the actual health behavior change sometimes try to achieve a target in a way in which it was not supposed to be achieved according to the game rules [13, 37]. Users cheat by, for example, "exploiting inherent sensor-related limitations to fabricate false detection" [37]. Cheating is also promoted by the fact that most gamified HBCSSs are used while being online and without supervision by professionals, which might lower the threshold for lying [55]. Cheating the self can lead to misdirected incentives and false health behavior and might thus negatively impact health-related outcomes [55].

Rewarding Incorrect Execution. Creating gamified HBCSSs requires a great variety of different resources and expertise [34]. In particular, developers need to ensure that games' rewards and progresses adequately reflect people's health behavior. If the gamification concept is not sufficiently aligned with the desired health behavior change, users might unintentionally be incentivized for wrong or even unhealthy behavior (e.g., distributing a reward although an exercise was not performed correctly) [36].

Overuse. Gamification frequently aims to motivate people to use an HBCSS in a more regular manner. However, gamification concepts that are not sufficiently balanced but instead reward exaggerated repetitions of certain tasks might incentivize users to use an HBCSS too excessively and thus overreaching their personal limits [55, 58]. In an extreme case, gamification concepts might lead to users that are "driven by obsession rather than enjoyment […] resulting in problems relating to overtraining, overexertion and risk taking" [57]. Barrat [57], for example, reported in his study about a gamified HBCSS for cycling that some participants – driven by peer pressure and competition – ended up cycling excessively thereby negatively influencing their overall health and social outcomes.

4.4 Demoralizing Effects

Cheating Others. As stated before, cheating can be an important problem in gamified HBCSSs. When users cheat they often aim to gain an advantage over other users by exploiting certain game mechanisms in a way that should not be allowed according to the game rules [37]. Cheating can ruin the fairness of competitive game mechanisms and might even result in other users giving up [37]. As a result, dissatisfaction among users that do not cheat might increase [13]. Recent research suggests that users are more likely to cheat the more they are exposed to other users cheating [62].

Overemphasized Peer Pressure. Gamification often contains competitive game elements that, for example, enable comparing the scores of different users [37]. However, competition and social comparison is not for everyone. Recent research suggests that introverted users are likely to be demotivated by gamification that contains social comparison [60]. In addition, competitive game elements that publicly compare the scores of different users can also be discouraging to those who have low scores due to temporary setbacks or simply being new to the system [37]. Thus, studies found that social comparison can have negative effects on users' health behavior [38]. For example, Horse-Fraile et al. [41] report of collaborative health app studies in which parents complained about the possibility that their children may become demotivated if other families ranked better than theirs did.

Exaggerated Punishment. Users might also be discouraged due to the feeling of disproportionate punishment. Even regularly active users might get sick or be otherwise unable to perform desired health-related behavior (e.g., exercises), resulting in a sharp drop of their (average) score [37]. Although this drop might sufficiently represent reality, users' might perceive it as unreasonable and unrepresentative of their health behavior and thus get discouraged or even stop using the HBCSS [37].

Feeling of Manipulation. Gamification aims to foster users' motivation to perform certain activities in a specific way or in a more regular manner, for example, by applying rule-based systems. This could cause users to perceive a feeling of being manipulated or forced into performing those actions, especially if the underlying health activity is inherently unstructured and requires a great degree of autonomy. For example, Barratt [57] reports that participants in a study of a cycling application complained about negative experiences due to a restricted level of autonomy.

Discouragement Due to Failure. Gamification often relies on goal-oriented game elements. These may discourage users if they fail to meet certain goals despite putting in a lot of effort. According to extant literature, it is particularly important for more serious contexts such as heart diseases [61] that developers avoid a sense of defeat while adjusting the level of difficulty according to users' capabilities.

4.5 Overstepping Boundaries

Privacy Infringements. Information about users' health status and health behavior is often sensitive and subject to specific laws which limit the disclosure of healthcare information without explicit consent from the user [36]. Implementing gamification in HBCSSs can add another level of complexity concerning privacy and data protection [13]. Based on game elements additional health information about HBCSS users might be gathered and stored. For example, users' badge collections might disclose information about health status and past health-related behaviors. This can be particularly problematic when unknown third parties, such as employers or insurance companies, start reviewing badge collections by individuals or user groups for specific purposes [36]. Research also indicates that some users are more likely to participate in HBCSSs if they are not required to disclose personal data [60].

Fostering Behavior that Harms Third Parties. Gamification of HBCSSs might also cause unintended side effects for third parties outside of the system. This was particularly observed in GPS-based HBCSSs that aim to foster physical activity and use auto-generated outdoor locations for specific rewards. For example, in 2011 an object used in a GPS-based gamified HBCSS was placed in the Downtown Disney Park in Anaheim, CA, USA. As a result, people that not played the game were scared and triggered a bomb alarm. This led to temporary closure of the park [58]. Another example for behavior that harms third parties might be overcrowding of specific real-world places that promise special rewards as observed in the beginning of Pokémon GO.

5 Discussion

Our objective within this research was the identification and analysis of unintended side effects of gamification in HBCSSs. Building on the results described in section four, we were generally able to provide answers to our research question and thus contributed to the knowledge base on gamification. We discuss some of the most interesting findings and their implications in the following.

By intensively reviewing and discussing the 16 identified side effects, we were able to build five logically coherent groups of unintended side effects. First, adverse motivational outcomes describe those side effects which basically result in the main objective of gamification (i.e., fostering user motivation and engagement) not being achieved. Reasons for the occurrence of side effects of this group are diverse. Many researchers emphasize that applying gamification is a demanding process that requires excessive resources and expertise in order to be effective [13, 15]. When developers decide not to include game design knowledge in the design process and instead go for "cheap" gamification solutions that are based on extrinsic rewards that are easy to implement and promise short-term behavior change, their gamification approach is likely to fail in the long run. If adverse motivational outcomes occur, they are likely to negatively impact users' adoption of gamified HBCSSs. Second, informational noise describes unintended side effects that are related to a flawed visual representation and interaction concept of the gamified HBCSSs. They are often caused by insufficient consideration of the unique seriousness of the health context and thus choosing inappropriate gamification elements. If those unintended side effects occur, users' might perceive an HBCSS as being less professional and serious. Third, reduced integrity of exercise describes those unintended side effects that can directly lead to health-related disadvantages for the user. Thus, it is particularly important to prevent these side effects from occurring. They are often caused by not carefully aligning the gamification concept of HBCSSs with their core activities and thus providing users with wrong incentives. Fourth, demoralization of users describes unintended side effects that might cause users to stop using the system because they feel treated unfairly. Fifth, overstepping of boundaries describes those side effects that might lead to legal issues for developers of gamified HBCSSs. In order to prevent legal uncertainties in particular when considering the high sensitivity of health-related data, it is very important for developers to prevent these side effects from occurring.

As mentioned before, some side effects are more specific to HBCSSs than others. For example, although cheating the self might occur in any gamification context, its potential consequences are particularly critical in HBCSSs as they threaten the correct execution of health behavior and thus yield potential harm to users' health. Another side effect that is strongly linked to HBCSSs is the trivialization of the context. While colorful and eye-catching design themes might be unproblematic for less serious contexts, within HBCSSs users might perceive them as especially inappropriate and unprofessional and subsequently avoid using the system for a serious context such as healthcare.

Our research yields some implications for practice. First, by presenting and discussing our results we show that developers of HBCSSs need to be aware of potential drawbacks of gamification. In addition, developers need a set of suitable prevention strategies to address these potential side effects when designing gamified HBCSSs. However, it was beyond the scope of this work to derive such prevention strategies. Thus, we leave it to future research and to developers of gamified HBCSSs to carefully develop prevention strategies that fit the diverse application scenarios of HBCSSs. Concerning implications for research, existing frameworks for gamification of HBCSSs mostly have been created on the basis of potential positive effects of gamification. With regard to our results we think that existing frameworks should be critically reviewed and evaluated whether they adequately consider unintended side effects of gamification.

Our research contributes to the scientific knowledge base in several aspects. By synthesizing knowledge on unintended side effects of gamification in HBCSSs, we contribute to the conceptual knowledge on gamification in the context of health and well-being. In particular, although we acknowledge the various potential positive aspects of gamification, we contribute to conceptual knowledge by taking on a more critical view of gamification, which has so far been underrepresented in gamification research [16, 29]. This critical view leads to a more nuanced perception of gamification and thus contributes to a better understanding of how to apply gamification as a valuable tool that motivates and does not harm users of HBCSSs.

The limitations of this study are as follows. First, our literature base is limited to 33 articles that discuss potential unintended side effects of gamification in HBCSSs. This shows that the topic received only little attention among researchers so far. Although we carefully performed our keyword-based search, we cannot rule out the possibility that we might have missed relevant articles. Second, there are different kinds of HBCSSs with various levels of context seriousness [2]. Thus, side effects might also differ between these different types of HBCSSs. It would be interesting for future research to delve deeper and compare potential side effects of different types of HBCSSs. Lastly, interpreting literature and text passages always leave some room for interpretation. However, we aimed to address this issue by carefully performing the concept-centric data analysis and developing a common understanding of side effects while synthesizing the literature. Future research might further strengthen the knowledge base by conducting qualitative research and gathering primary data from experts of the field. Future research could also aim to delve deeper in order to understand the underlying reasons for the observed side effects and the circumstances under which they occur.

6 Conclusion

Although gamification has received tremendous attention from practitioners and researchers interested in HBCSSs, little attention has been paid to the unintended side effects that potentially come with the implementation of gamification in HBCSSs. Within this research, we have taken a first step towards closing this gap by reviewing past research on gamification in HBCSSs and synthesizing 16 unintended side effects that may occur when using gamification. Our research adds to the knowledge base on gamification by, for the first time, providing a comprehensive overview of potential unintended side effects. We also contribute to a more nuanced view of gamification, which helps relieving gamification from unrealistic expectations threatening to position it as a silver bullet rather than a valuable tool for developers of HBCSSs. Future research should delve deeper into this topic and examine in more detail under which circumstances such side effects can occur and how they can be avoided or counteracted.

7 Acknowledgements

This study was partially supported by the Swiss National Science Foundation (grant no.: 172740).

References

1. Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., Hides, L.: Gamification for health and wellbeing: A systematic review of the literature. Internet interventions 6, 89-106 (2016)
2. Stepanovic, S., Mettler, T.: Gamification applied for health promotion: does it really foster long-term engagement? A scoping review. European Conference on Information Systems, Portsmouth, UK (2018)
3. Oinas-Kukkonen, H.: A foundation for the study of behavior change support systems. Personal and ubiquitous computing 17, 1223-1235 (2013)
4. Alahäivälä, T., Oinas-Kukkonen, H.: Understanding persuasion contexts in health gamification: a systematic analysis of gamified health behavior change support systems literature. International Journal of Medical Informatics 96, 62-70 (2016)
5. Jones, B.A., Madden, G.J., Wengreen, H.J.: The FIT Game: preliminary evaluation of a gamification approach to increasing fruit and vegetable consumption in school. Preventive Medicine 68, 76-79 (2014)
6. Jones, B.A., Madden, G.J., Wengreen, H.J., Aguilar, S.S., Desjardins, E.A.: Gamification of dietary decision-making in an elementary-school cafeteria. PLoS One 9, e93872 (2014)
7. Chung, A.E., Skinner, A.C., Hasty, S.E., Perrin, E.M.: Tweeting to health: A novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clinical Pediatrics 56, 26-32 (2017)
8. Hamari, J., Koivisto, J.: 'Working out for likes': An empirical study on social influence in exercise gamification. Computers in Human Behavior 50, 333-347 (2015)
9. Stinson, J.N., Jibb, L.A., Nguyen, C., Nathan, P.C., Maloney, A.M., Dupuis, L.L., Gerstle, J.T., Alman, B., Hopyan, S., Strahlendorf, C.: Development and testing of a

multidimensional iPhone pain assessment application for adolescents with cancer. Journal of Medical Internet Research 15, e51 (2013)
10. AlMarshedi, A., Wills, G., Ranchhod, A.: Gamifying self-management of chronic illnesses: a mixed-methods study. JMIR serious games 4, (2016)
11. Krebs, P., Duncan, D.T.: Health app use among US mobile phone owners: A national survey. JMIR mHealth and uHealth 3, e101 (2015)
12. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining gamification. Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, pp. 9-15. Tampere, Finland (2011)
13. Pereira, P., Duarte, E., Rebelo, F., Noriega, P.: A review of gamification for health-related contexts. International conference of design, user experience, and usability, pp. 742-753. Springer (2014)
14. Esteves, J.M.: The perils of gamification triviliazation: how and why gamification is failing to deliver loyalty. Academy of Management Proceedings 2017, 13560 (2017)
15. Sardi, L., Idri, A., Fernández-Alemán, J.L.: A Systematic Review of Gamification in e-Health. Journal of Biomedical Informatics 71, 31-48 (2017)
16. Hyrynsalmi, S., Smed, J., Kimppa, K.K.: The Dark Side of Gamification: How We Should Stop Worrying and Study also the Negative Impacts of Bringing Game Design Elements to Everywhere. Proceedings of the 1st International GamiFIN Conference, pp. 105-110, Tampere, Finland (2017)
17. Toda, A.M., Valle, P.H., Isotani, S.: The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education. Researcher Links Workshop: Higher Education for All, pp. 143-156. Springer (2017)
18. Matallaoui, A., Koivisto, J., Hamari, J., Zarnekow, R.: How Effective Is 'Exergamification'? A Systematic Review on the Effectiveness of Gamification Features in Exergames. Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 3316-3325, Big Island, Hawaii, United States (2017)
19. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work?--a literature review of empirical studies on gamification. 47th Hawaii International Conference on System Sciences (HICSS), 2014 pp. 3025-3034. IEEE, Waikoloa, HI (2014)
20. AlMarshedi, A., Wills, G.B., Ranchhod, A.: The Wheel of Sukr: a framework for gamifying diabetes self-management in Saudi Arabia. Procedia Computer Science 63, 475-480 (2015)
21. Thiebes, S., Lins, S., Basten, D.: Gamifying Information Systems-A Synthesis of Gamification Mechanics and Dynamics. European Conference on Information Systems, pp. 1-17, Tel Aviv, Israel (2014)
22. Huotari, K., Hamari, J.: Defining gamification: a service marketing perspective. Proceedings of the 16th International Academic MindTrek Conference, pp. 17-22. ACM, Tampere, Finland (2012)
23. McGonigal, J.: Reality is broken: Why games make us better and how they can change the world. Penguin (2011)
24. Fleming, T.M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S.N., Cheek, C., Lucassen, M., Lau, H.M., Pollmuller, B., Hetrick, S.: Serious games and gamification for mental health: current status and promising directions. Frontiers in psychiatry 7, 215 (2017)
25. Putz, L.-M., Schmidt-Kraepelin, M., Treiblmaier, H., Sunyaev, A.: The influence of gamified workshops on students' knowledge retention. Proceedings of the 2nd International GamiFIN Conference, Tampere, Finland (2018)
26. Nah, F.F.-H., Zeng, Q., Telaprolu, V.R., Ayyappa, A.P., Eschenbrenner, B.: Gamification of education: a review of literature. International conference on hci in business, pp. 401-409. Springer (2014)

27. Morschheuser, B., Hamari, J., Koivisto, J.: Gamification in crowdsourcing: a review. 49th Hawaii International Conference on System Sciences, pp. 4375-4384. IEEE (2016)
28. Schmidt-Kraepelin, M., Thiebes, S., Tran, M.C., Sunyaev, A.: What's in the Game? Developing a Taxonomy of Gamification Concepts for Health Apps. Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 1217-1226, Waikoloa, Hawaii, USA (2018)
29. Kim, T.W., Werbach, K.: More than just a game: ethical issues in gamification. Ethics and Information Technology 18, 157-173 (2016)
30. Levy, Y., Ellis, T.J.: A systems approach to conduct an effective literature review in support of information systems research. Informing Science 9, (2006)
31. Strauss, A., Corbin, J.M.: Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications, Inc (1990)
32. Hamari, J., Koivisto, J.: Why do people use gamification services? International Journal of Information Management 35, 419-431 (2015)
33. Attig, C., Franke, T.: I track, therefore I walk–Exploring the motivational costs of wearing activity trackers in actual users. International Journal of Human-Computer Studies (2018)
34. Forman, E.M., Goldstein, S.P., Flack, D., Evans, B.C., Manasse, S.M., Dochat, C.: Promising technological innovations in cognitive training to treat eating-related behavior. Appetite 10, (2017)
35. Mitchell, R., Schuster, L., Drennan, J.: Understanding how gamification influences behaviour in social marketing. Australasian Marketing Journal (AMJ) 25, 12-19 (2017)
36. McDaniel, R.: A taxonomy for digital badge design in medical technologies. 2016 IEEE International Conference on Serious Games and Applications for Health, pp. 1-8 (2016)
37. Gal-Oz, A., Zuckerman, O.: Embracing Cheating in Gamified Fitness Applications. Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, pp. 535-540. ACM, London, United Kingdom (2015)
38. Zuckerman, O., Gal-Oz, A.: Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Personal and ubiquitous computing 18, 1705-1719 (2014)
39. Helf, C., Hlavacs, H.: Apps for life change: Critical review and solution directions. Entertainment Computing 14, 17-22 (2016)
40. El-Hilly, A.A., Iqbal, S.S., Ahmed, M., Sherwani, Y., Muntasir, M., Siddiqui, S., Al-Fagih, Z., Usmani, O., Eisingerich, B.A.: Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study. JMIR Serious Games 4, e18 (2016)
41. Hors-Fraile, S., Atique, S., Mayer, M., Denecke, K., Merolli, M., Househ, M.: The unintended consequences of social media in healthcare: new problems and new solutions. Yearbook of medical informatics 25, 47-52 (2016)
42. Lewis, Z.H., Swartz, M.C., Lyons, E.J.: What's the point?: A review of reward systems implemented in gamification interventions. Games for health journal 5, 93-99 (2016)
43. Zhao, Z., Arya, A., Whitehead, A., Chan, G., Etemad, S.A.: Keeping Users Engaged through Feature Updates: A Long-Term Study of Using Wearable-Based Exergames. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1053-1064. ACM (2017)
44. AlMarshedi, A., Wills, G., Ranchhod, A.: Guidelines for the Gamification of Self-Management of Chronic Illnesses: Multimethod Study. JMIR serious games 5, (2017)
45. Boendermaker, W.J., Maceiras, S.S., Boffo, M., Wiers, R.W.: Attentional bias modification with serious game elements: evaluating the shots game. JMIR serious games 4, (2016)

46. Gremaud, A.L., Carr, L.J., Simmering, J.E., Evans, N.J., Cremer, J.F., Segre, A.M., Polgreen, L.A., Polgreen, P.M.: Gamifying Accelerometer Use Increases Physical Activity Levels of Sedentary Office Workers. Journal of the American Heart Association 7, (2018)
47. Kouwenhoven-Pasmooij, T.A., Robroek, S.J., Ling, S.W., van Rosmalen, J., van Rossum, E.F., Burdorf, A., Hunink, M.M.: A Blended Web-Based Gaming Intervention on Changes in Physical Activity for Overweight and Obese Employees: Influence and Usage in an Experimental Pilot Study. JMIR serious games 5, (2017)
48. Patel, M.S., Benjamin, E.J., Volpp, K.G., Fox, C.S., Small, D.S., Massaro, J.M., Lee, J.J., Hilbert, V., Valentino, M., Taylor, D.H.: Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: The BE FIT randomized clinical trial. JAMA internal medicine 177, 1586-1593 (2017)
49. Lumsden, J., Skinner, A., Coyle, D., Lawrence, N., Munafo, M.: Attrition from Web-Based Cognitive Testing: A Repeated Measures Comparison of Gamification Techniques. Journal of medical Internet research 19, (2017)
50. Zhang, M., Ying, J., Song, G., Fung, D.S., Smith, H.: Gamified Cognitive Bias Modification Interventions for Psychiatric Disorders. JMIR mental health 5, e11640 (2018)
51. Lin, Y., Tudor-Sfetea, C., Siddiqui, S., Sherwani, Y., Ahmed, M., Eisingerich, A.B.: Effective Behavioral Changes through a Digital mHealth App: Exploring the Impact of Hedonic Well-Being, Psychological Empowerment and Inspiration. JMIR mHealth and uHealth 6, e10024 (2018)
52. Lumsden, J., Edwards, E.A., Lawrence, N.S., Coyle, D., Munafò, M.R.: Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR serious games 4, (2016)
53. Brown, M., O'Neill, N., van Woerden, H., Eslambolchilar, P., Jones, M., John, A.: Gamification and adherence to Web-based mental health interventions: a systematic review. JMIR mental health 3, (2016)
54. Lister, C., West, J.H., Cannon, B., Sax, T., Brodegard, D.: Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2, e9 (2014)
55. Hopia, H., Raitio, K.: Gamification in Healthcare: Perspectives of Mental Health Service Users and Health Professionals. Issues in mental health nursing 37, 894-902 (2016)
56. Miller, A.S., Cafazzo, J.A., Seto, E.: A game plan: Gamification design principles in mHealth applications for chronic disease management. Health Informatics Journal 22, 184-193 (2016)
57. Barratt, P.: Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling. Health & place (2016)
58. Boulos, M.N.K., Yang, S.P.: Exergames for health and fitness: the roles of GPS and geosocial apps. BioMed Central (2013)
59. van Mierlo, T., Hyatt, D., Ching, A.T., Fournier, R., Dembo, R.S.: Behavioral economics, wearable devices, and cooperative games: results from a population-based intervention to increase physical activity. JMIR serious games 4, (2016)
60. Orji, R., Nacke, L.E., Di Marco, C.: Towards personality-driven persuasive health games and gamified systems. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1015-1027. ACM (2017)
61. Dithmer, M., Rasmussen, J.O., Grönvall, E., Spindler, H., Hansen, J., Nielsen, G., Sørensen, S.B., Dinesen, B.: "The Heart Game": Using Gamification as Part of a Telerehabilitation Program for Heart Patients. Games for Health Journal 5, 27-33 (2016)
62. Wu, Y., Chen, V.H.H.: A social-cognitive approach to online game cheating. Computers in Human Behavior 29, 2557-2567 (2013)

Investigating the Influence of Information Incongruity on Trust-Relations within Trilateral Healthcare Settings

Marius Mueller[1], Oliver Heger[1], Bastian Kordyaka[1], and Bjoern Niehaves[1]

[1] University of Siegen, Chair of Information Systems, Siegen, Germany
{firstname.lastname}@uni-siegen.de

Abstract. Modern health information technologies (HIT) come with many benefits for healthcare, such as a decrease of necessary clinical visits or independent health monitoring. The deployment of these technologies to support medical treatments expands the traditional patient-physician relationship to a trilateral setting involving patient, physician, and HIT. Whereas patients formerly relied on health-related information given by their physician, the digitization of healthcare as well as increasing levels of individual health literacy represent new sources of information and, thus, call for investigating different forms of trust towards medical experts, technologies, and the patient's own judgements. Information incongruities, however, can lead to new forms of trust issues, thus calling for dedicated research. We propose a vignette study in the form of an online survey to investigate the influence information incongruities can have on different forms of patient-sided trust. For this, we develop hypotheses representing our expected results.

Keywords: HIT, Trust, Information Incongruity, Health Literacy, Self-efficacy.

1 Introduction

Traditional healthcare settings involve relationships between patients and one or more medical experts such as primary care physicians that are largely based on interpersonal trust, empathy, and satisfaction with treatments [1]. Patient-sided trust in physicians leads to compliance with instructions, assessments, and advices [2], which in turn increases the effectiveness of medical treatments. Whereas traditional patient-physician relationships evolve around the patient following and complying with recommendations and judgements made by the physician, a deliberative relationship model emerged throughout the last decades that calls for a higher degree of patient participation and autonomy regarding clinical processes [3, 4]. This autonomous and deliberative stance becomes more important in the light of modern developments and deployments of technologies for healthcare, often referred to as health IT (HIT) [5].

The implementation of HIT within medical processes and, thus, patient-physician relationships introduces a new source of health-related information. Hence, the medical setting becomes more complex and new relationships are formed. On the one hand, patients interact with mobile, self-managed HIT systems [6] or more complex ones that are deployed in their home environment [7, 8]. For example, they are enabled to

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

autonomously retrieve information about their health status and the potential need for interventions. Wearable sensors, for instance, can measure the patient's heart rate and cardiac status throughout the day [9]. On the other hand, physicians deploy HIT systems to support patient care, to reduce the number of face-to-face meetings through health status monitoring (gathered by wearable sensors or patient input), or to mediate therapeutic instructions using telemedical tools such as live video sessions [10]. Hence, HIT represents a new 'actor' within healthcare settings, enabling new forms of interaction. A transition from bilateral to trilateral relationships becomes apparent.

However, by dissolving former power-imbalances between patients and physicians through HIT deployment, these new forms of relationships are prone to deteriorations, for instance evoked by informational gaps and differing outcome expectations between patients, experts, and technologies [11]. Since HIT systems represent an additional source of health-related information, new issues arise. The HIT's behaviour and output might contradict the information the patient or physician have, which potentially evokes questions on whether the patient or physician might be wrong in their assessment [11]. This, in turn, might deteriorate prevalent trust-relations. From a patient perspective, three trust-forming relations can be identified, which are trust in the physician [1], trust in the HIT [12], and trust in oneself (regarding knowledge and actions) [13]. A decrease in trust can originate from different forms of information incongruity and resulting skepticism towards the source of information trusted the least [12]. As suggested by Cognitive Balance Theory, imbalances within a trilateral setting (e.g. patient trusts physician and HIT but physician and HIT contradict each other) lead to discomfort [14].

This study's goal is to shed light on the emergence of trust issues within trilateral healthcare settings evoked by information incongruity between involved actors, which delivers implications for HIT design and solutions to hamper the deterioration of trust. Consequently, this study is guided by the following research question (RQ):

RQ: *How do information incongruities within a trilateral healthcare setting influence patient-sided forms of trust (trust in physician, a HIT, and in oneself)?*

2 Trust in Physician, HIT, and Oneself

Trust is the willingness of people to be exposed in risky situations [15] and depend on another party [16]. From the patient's perspective within the trilateral setting, trust is their willingness to depend their health on 1) the physician, 2) the health information technology, and 3) their own health knowledge and capabilities.

Trust between a patient and a physician depends on many factors. Among others, significant drivers of interpersonal trust within a patient-physician relationship are perceived empathy, the patient's satisfaction with courses and outcomes of treatments [1], patient-centered communication [17], and autonomy [18] as well as perceived control [19]. Patient-sided trust in recommendations, treatment instructions, and overall assessments made by a physician is expressed through the patient's compliance [2]. Non-compliance can interfere with the patient's health and therapeutic efforts [20]. As a result, the effectiveness and outcome of therapies and medical treatments greatly depend on this relationship and emerging trust-relations.

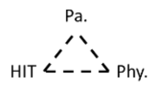
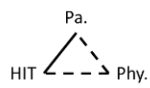
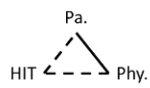
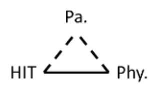
Besides trust in the human expert, trust in HIT itself plays a role in shaping the trilateral setting. Trusting beliefs in a specific technology is shaped by trust in its functionality, reliability, and helpfulness [21]. The formation of trusting beliefs in IT can be explained by ease of use, system quality perceptions, uncertainty avoidance culture, and institution-based trust [22]. Furthermore, the formation of trust in technology is influenced by the performance, process, and purpose of the IT artifact itself [23]. Trust is not only a desirable condition within a medical setting itself, but an important precondition for a successful adoption of "risky" and novel technologies [24].

The third trust relationship of interest is the patients' trust in their own competence to assess the appropriateness of a medical treatment and/or to carry out the medical treatment by themselves (e.g. physiotherapeutic activities at home). This trust relationship can be represented by the constructs of 'self-efficacy' and 'health literacy'. Self-efficacy deals with one's "*judgment of their capabilities to organize and execute courses of action required to attain designated types of performance*" [25] (p. 391). Studies indicate that self-efficacy related to health practices shows strong relationships with health behavior [13]. Besides this, computer self-efficacy could additionally play a significant role within the trilateral setting, since the patient is expected to use technology [26]. Similarly but from a more dispositional perspective, health literacy is "*the ability to understand and interpret the meaning of health information in written, spoken or digital form*" [27] (p. 144). Health literacy is an essential factor when it comes to the deployment of HIT, since their success depends on the effective communication with different audiences, who have unique needs and capacities [28].

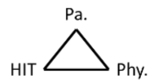
3 Hypotheses

Based on the trilateral setting including patient, HIT, and physician, we derived the following hypotheses (Table 1):

Table 1. Hypotheses (Pa.=Patient, Phy.=Physician; solid=congruent, dashed=incongruent)

Hypothesis	*Illustration*
H1: Information incongruity between patient, HIT and physician negatively influences trust in patient, HIT, and physician.	
H2a: Information congruity between patient and HIT and information incongruity between patient or HIT and physician negatively influences trust in physician.	
H2b: Information congruity between patient and physician and information incongruity between patient or physician and HIT negatively influences trust in HIT.	
H2c: Information congruity between HIT and physician and information incongruity between HIT or physician and patient negatively influences trust in patient (self-efficacy).	

H3: Information congruity between patient, HIT and physician positively influences trust in patient, HIT, and physician.



4 Method and Outlook

To empirically test our hypotheses, we plan to conduct a between-subjects vignette study with an online survey given out to a self-selected convenient sample consisting of patients. To acquire a sample of sufficient size, we distribute the survey link via a network of regional primary care physicians as well as via social media. Each of the five conditions within the trilateral setting (cf. Table 1) is represented by two vignettes. To ensure external validity, we interview primary care physicians in a preliminary step. In doing so, we aim for constructing viable vignettes that depict common treatment scenarios that often occur among patients with chronic conditions. Since trust forms over time, the investigation of such conditions seems promising. Hence, we target participants who have experienced chronic treatments.

We aim for constructing two vignettes. The first vignette describes a scenario in which the participant is instructed by a HIT to take pills and receives congruent (or incongruent) information. The second vignette describes a more complex scenario in which the participant has to carry out a physiotherapeutic exercise. Following each vignette, the participant has to answer the same questionnaire. The two vignettes are presented in a random order. The reasons for presenting two vignettes per condition is to account for external validity and to alter the degree of scenario complexity, since trust is especially relevant in complex situations [29].

For the questionnaire, we plan to adapt the eleven items of the "Trust in Physician Scale" [19] to measure interpersonal trust between participants and the physician. For the participants' trust in HIT, we adapt the eleven "Trust Belief in Specific Technology" items from McKnight et al. [21]. For participants' trust in themselves, we adapt relevant items from the SRAHP scale [30] and the European Health Literacy Questionnaire (HLS-EU-Q) [31]. We further include demographic and control variables. For data analysis, we plan to conduct three separate 2 (simple vs. complex scenario) x 5 (trilateral trust settings depicted in Table 1) mixed ANOVAs with complexity as within-subject factor, information congruity as between-subject factor, and the three forms of trust (trust in physician, the HIT, and in oneself) as dependent variables.

Based on the results of this study, we plan to derive and empirically evaluate technology design implications to deal with the matter of trust. Although the trilateral setting described in this paper is a simplified model, we believe that considering potential information incongruities in the design of health information technologies will increase trust in and the acceptance of digital healthcare interventions.

5 Acknowledgements

This research is supported by the SenseVojta project granted by the German Federal Ministry of Education and Research (BMBF) (No. 13GW0166E).

References

1. Hojat, M., Louis, D.Z., Maxwell, K., Markham, F., Wender, R., Gonnella, J.S.: Patient perceptions of physician empathy, satisfaction with physician, interpersonal trust, and compliance. International Journal of Medical Education. 1, 83–87 (2010)
2. Kim, S.S., Kaplowitz, S., Johnston, M.V.: The Effects of Physician Empathy on Patient Satisfaction and Compliance. Evaluation & the Health Professions. 27, 237–251 (2004)
3. Brody, D.S.: The patient's role in clinical decision-making. Annals of Internal Medicine. 93, 718–722 (1980)
4. Emanuel, E.J., Emanuel, L.L.: Four models of the physician-patient relationship. JAMA. 267, 2221–2226 (1992)
5. Goldschmidt, P.G.: HIT and MIS: implications of health information technology and medical information systems. Communications of the ACM. 48, 68–74 (2005)
6. Cafazzo, J.A., Casselman, M., Hamming, N., Katzman, D.K., Palmert, M.R.: Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. Journal of Medical Internet Research. 14, e70 (2012)
7. Tang, R., Yang, X.-D., Bateman, S., Jorge, J., Tang, A.: Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 4123–4132. ACM Press, Seoul, Republic of Korea (2015)
8. Tetteroo, D., Vreugdenhil, P., Grisel, I., Michielsen, M., Kuppens, E., Vanmulken, D., Markopoulos, P.: Lessons Learnt from Deploying an End-User Development Platform for Physical Rehabilitation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 4133–4142. ACM Press, Seoul, Republic of Korea (2015)
9. Pantelopoulos, A., Bourbakis, N.G.: A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 40, 1–12 (2010)
10. Kvedar, J., Coye, M.J., Everett, W.: Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Affairs. 33, 194–199 (2014)
11. Van Woerkum, C.M.: The Internet and primary care physicians: coping with different expectations. The American Journal of Clinical Nutrition. 77, 1016–1018 (2003)
12. Hesse, B.W., Nelson, D.E., Kreps, G.L., Croyle, R.T., Arora, N.K., Rimer, B.K., Viswanath, K.: Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Archives of internal medicine. 165, 2618–2624 (2005)
13. Strecher, V.J., McEvoy DeVellis, B., Becker, M.H., Rosenstock, I.M.: The role of self-efficacy in achieving health behavior change. Health education quarterly. 13, 73–92 (1986)
14. Heider, F.: The psychology of interpersonal relations. John Wiley & Sons Inc., Hoboken, NJ, US (1958)
15. Gefen, D., Straub, D.W.: The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption. Journal of the Association for Information Systems. 1, 8–37 (2000)
16. Rousseau, D.M., Sitkin, S.B., Burt, R.S., Camerer, C.: Not so different after all: A cross-discipline view of trust. Academy of management review. 23, 393–404 (1998)
17. Epstein, R.M., Franks, P., Fiscella, K., Shields, C.G., Meldrum, S.C., Kravitz, R.L., Duberstein, P.R.: Measuring patient-centered communication in patient–physician

consultations: theoretical and practical issues. Social science & medicine. 61, 1516–1528 (2005)
18. Lee, Y.-Y., Lin, J.L.: Do patient autonomy preferences matter? Linking patient-centered care to patient–physician relationships and health outcomes. Social science & medicine. 71, 1811–1818 (2010)
19. Anderson, L.A., Dedrick, R.F.: Development of the Trust in Physician scale: a measure to assess interpersonal trust in patient-physician relationships. Psychological reports. 67, 1091–1100 (1990)
20. Roter, D.L., Hall, J.A., Merisca, R., Nordstrom, B., Cretin, D., Svarstad, B.: Effectiveness of interventions to improve patient compliance: a meta-analysis. Medical care. 36, 1138–1161 (1998)
21. Mcknight, D.H., Carter, M., Thatcher, J.B., Clay, P.F.: Trust in a specific technology: An investigation of its components and measures. ACM Transactions on Management Information Systems (TMIS). 2, 12 (2011)
22. Vance, A., Elie-Dit-Cosaque, C., Straub, D.W.: Examining trust in information technology artifacts: the effects of system quality and culture. Journal of management information systems. 24, 73–100 (2008)
23. Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., Leimeister, J.M.: Understanding the formation of trust in IT artifacts. In: Proceedings of the 33rd International Conference on Information Systems. , Orlando, USA (2012)
24. Fukuyama, F.: Trust: The social virtues and the creation of prosperity. Free Press Paperbacks, NY (1995)
25. Bandura, A.: Social foundations of thought and action. Englewood Cliffs, NJ. 1986, (1986)
26. Compeau, D.R., Higgins, C.A.: Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly. 19, 189–211 (1995)
27. Adams, R.J., Stocks, N.P., Wilson, D.H., Hill, C.L., Gravier, S., Kickbusch, L., Beilby, J.J.: Health literacy-a new concept for general practice? Australian family physician. 38, 144–147 (2009)
28. Kreps, G.L.: The relevance of health literacy to mHealth. Information Services & Use. 37, 123–130 (2017)
29. Luhmann, N.: Trust and power. John Wiley and Sons, Inc., Chichester, UK (1979)
30. Becker, H., Stuifbergen, A., Oh, H.S., Hall, S.: Self-rated abilities for health practices: A health self-efficacy measure. Health Values: The Journal of Health Behavior, Education & Promotion. 17, 42–50 (1993)
31. Sørensen, K., Van den Broucke, S., Pelikan, J.M., Fullam, J., Doyle, G., Slonska, Z., Kondilis, B., Stoffels, V., Osborne, R.H., Brand, H.: Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health. 13, 948 (2013)

Track 9:

Krisen- und Kontinuitätsmanagement

Track Chairs:

Prof. Dr. Christian Reuter
Technische Universität Darmstadt

Prof. Dr. Stefan Stieglitz
University of Duisburg-Essen

Dr. Marén Schorch
Universität Siegen

Potentiale von IKT beim Ausfall kritischer Infrastrukturen: Erwartungen, Informationsgewinnung und Mediennutzung der Zivilbevölkerung in Deutschland

Marc-André Kaufhold[1,2], Margarita Grinko[1], Christian Reuter[1,2], Marén Schorch[1], Amanda Langer[1], Sascha Skudelny[1], and Matthias Hollick[3]

[1]Universität Siegen, Institut für Wirtschaftsinformatik, KontiKat
{marc.kaufhold@, margarita.grinko@student, maren.schorch@, amanda.langer@, sascha.skudelny@}uni-siegen.de
[2]Technische Universität Darmstadt,
Wissenschaft und Technik für Frieden und Sicherheit (PEASEC)
{reuter, kaufhold}@peasec.tu-darmstadt.de
[3]Technische Universität Darmstadt, Sichere Mobile Netze (SEEMOO)
matthias.hollick@seemoo.tu-darmstadt.de

Abstract. In der Sicherheits- und Krisenforschung stehen bislang primär Fragen der Vulnerabilität, Stärkung von Resilienz und Erhaltung bzw. Wiederherstellung kritischer Infrastrukturen (KRITIS) im Mittelpunkt; zunehmend wird auch die Bedeutung von Sozialen Medien und Krisen-Apps erkannt. Inwiefern ist die Zivilbevölkerung in Deutschland aber auf das Eintreten einer Krise tatsächlich vorbereitet? Welche Informations- und Kommunikationstechniken (IKT) werden im Alltag und bei einem potentiellen Infrastrukturausfall genutzt? Unser Beitrag stellt die Ergebnisse einer repräsentativen Umfrage mit 1024 Teilnehmern in Deutschland vor, die belegen, dass Gefahrenbewusstsein, Vorbereitung, effektives Krisenmanagement und Verbreitung von Krisen-Apps in Deutschland noch relativ gering sind, während traditionelle Kommunikations- und Informationskanäle sowie informelle Informationsnetzwerke bevorzugt werden. Die Ergebnisse stellen weiterhin die beträchtlichen Unterstützungspotentiale von IKT in derartigen Krisensituationen heraus, die zur Erhöhung des Risikobewusstseins, der Erleichterung des Informationstransfers und der Verbesserung der Kommunikation zwischen Zivilbevölkerung, KRITIS-Betreibern und Behörden und Organisationen mit Sicherheitsaufgaben (BOS) entscheidend beitragen können.

Keywords: Kritische Infrastrukturen, Informations- und Kommunikationstechnologien, Krisenkommunikation, Krisenmanagement, repräsentative Studie

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

1 Einleitung

Obgleich in Europa extreme Naturereignisse wie Hochwasser, Überschwemmungen und Stürme besonders häufig vorkommen [1], ist Deutschland aufgrund seiner Infrastruktur indes eher selten von kollektiven Krisen[1] und längerfristigen Infrastrukturausfällen von z. B. Strom und Telekommunikation betroffen [2]. Ein geringes Gefahrenbewusstsein der Zivilbevölkerung erhöht jedoch das Schadenspotential, denn das Gefühl für die eigene, potentielle Verwundbarkeit und eine entsprechende Prävention für Ausfälle scheint nicht präsent zu sein [3]. Im Krisenfall greifen die betroffenen Parteien auf verschiedene Quellen zurück, um sich Informationen zu beschaffen und diese anzubieten, die zivilgesellschaftliche Kontinuität zu wahren und unter anderem mittels Informations- und Kommunikationstechnologien (IKT) zusammenzuarbeiten [4], [5]. Während der IKT-Einsatz in Krisen vor allem seit den Anschlägen vom 11. September 2001 in New York ein eigenes Forschungsgebiet – Crisis Informatics – darstellt [4], fehlt es noch an repräsentativen Studien zu Informationserwartungen, Vorbereitungen für und Verhalten während Krisen – insbesondere in Mitteleuropa. Unsere Studie versucht, dieses Desiderat zu adressieren und beschäftigt sich mit folgenden vier Forschungsfragen:

FF1: Wie schätzen Bürger in Deutschland ihr eigenes Verhalten sowie den Grad an Vorbereitungen und Wissen bei einem Infrastrukturausfall ein?

FF2: Über welche Medienkanäle kommunizieren und informieren sie sich in Krisensituationen und warum?

FF3: Welche Arten und Kanäle von Informationen erwarten sie von KRITIS-Betreibern?

FF4: Wie kann IKT das Bewusstsein schärfen, Verhalten anleiten sowie die Kommunikation und zivilgesellschaftliche Kontinuität in solchen Situationen unterstützen?

Dazu wird in unserem Beitrag zunächst der bisherige Forschungsstand zum soziotechnischen Charakter kritischer Infrastrukturen (KRITIS) und der Krisenkommunikation analysiert (Abschnitt 2). Im Anschluss werden die Methodik (Abschnitt 3) und die Ergebnisse einer repräsentativen Umfrage in Deutschland (Abschnitt 4) geschildert, die auf die Beantwortung der zuvor gestellten Forschungsfragen zielen. In der anschließenden Diskussion und im Fazit (Abschnitt 5) werden die Ergebnisse hinsichtlich der vier Forschungsfragen reflektiert und die Implikationen für weitere Forschung und IKT-Gestaltung zusammengefasst.

Die Ergebnisse deuten auf ein geringes Bewusstsein und eine geringe Vorbereitung in Bezug auf das Risiko eines KRITIS-Ausfalls hin. Informationsbeschaffung und Kommunikation über verschiedene Medienkanäle erzeugen einen hohen Bedarf an spezifischen, zeitnahen und zuverlässigen Informationen und Handlungsempfehlungen. Zudem stellt dieser Beitrag die Potentiale der IKT-Nutzung in KRITIS-Ausfällen und Anforderungen an deren Gestaltung in Bezug auf Bedürfnisse und Eigenschaften der Nutzer heraus.

---

[1] Unter einer kollektiven Krise wird eine öffentlich wahrgenommene, beschleunigt (rapide) in Erscheinung tretende gravierende Problemsituation oder -entwicklung, die mit den üblichen Problemlösungsverfahren nicht bewältigt werden kann, verstanden [27].

2 Stand der Forschung

Eigenschaften sozio-technischer kritischer Infrastruktur. Infrastruktur kann definiert werden als ein zugrunde liegender Rahmen, welcher einer Gruppe, Organisation oder der Gesellschaft allgemein ein Funktionieren ermöglicht und aus materiellen, institutionellen und persönlichen Teilaspekten besteht [6]. In diesem Fall wird die materielle Kategorie betrachtet, vor allem die Bereitstellung von Ressourcen wie Energie und (Tele-)Kommunikation, die auch als kritische Infrastrukturen (KRITIS) bezeichnet werden. Der Rat der Europäischen Union definiert KRITIS als eine „Anlage, ein System oder ein Teil davon, die von wesentlicher Bedeutung für die Aufrechterhaltung wichtiger gesellschaftlicher Funktionen, der Gesundheit, der Sicherheit und des wirtschaftlichen oder sozialen Wohlergehens der Bevölkerung sind und deren Störung oder Zerstörung erhebliche Auswirkungen […] hätte" [7], in Form einer Unterbrechung der gesellschaftlichen und betrieblichen Kontinuität [8].

Auch wenn Infrastrukturausfälle nicht völlig vermeidbar sind, ist die Stromversorgung in Deutschland vergleichsweise stabil; Stromausfälle sind selten und relativ kurz [2]. Dies hat zur Folge, dass viele Bürger in Deutschland noch nie einen solchen Notfall erlebt haben und dementsprechend unvorbereitet sind, z. B. in Bezug auf Risiken und Sicherheitsmaßnahmen sowie Kommunikations- und Informationskanäle. Die Bürger erwarten, dass sich Behörden und Organisationen mit Sicherheitsaufgaben (BOS) um die potenzielle Krise kümmern [9]. Selbst wenn die Bürger eine Krise erlebt haben, ist die Wahrscheinlichkeit, dass sie vorbeugende Maßnahmen ergreifen, recht gering, da sie ein erneutes Auftreten für nicht wahrscheinlich erachten.

Die Vorbereitung auf eine Krise und die Stabilität einer KRITIS kann das Risiko von Zusammenbrüchen jedoch nur verringern, sodass die Reaktionen der Bürger und Behörden entscheidend für das Notfallmanagement, die Systemkontinuität und den Wiederaufbau sind. Diese „soziale Resilienz" beinhaltet allerdings nicht nur die Vorbereitung der Bürger auf eine mögliche Krise, sondern auch das Ermöglichen einer effektiven Ausbildung, Rahmenvereinbarungen und Kommunikation zwischen Experten, Behörden, Infrastrukturbetreibern, Medien, Einsatzkräften, Freiwilligen und Bürgern [8]. Hier entsteht „menschliche" oder „soziale Infrastruktur" [10], d.h. Organisationen, Teams oder Gruppen von Menschen, die gemeinsam agieren, sich unterstützen, kommunizieren und Informationen austauschen. Heutzutage werden diese Zusammenschlüsse zunehmend durch die Benutzung von sozialen Medien etabliert und tragen dazu bei, bestimmte Infrastrukturfunktionen und öffentliche Sicherheit aufrechtzuerhalten und effizienter wiederherzustellen [11].

Krisenkommunikation und Voraussetzungen für IKT. Krisenkommunikation bezieht sich meist auf die One-to-Many-Kommunikation und das öffentliche Image einer Organisation [12] – in unserem Fall das des KRITIS-Betreibers. Zusammen mit offiziellen Informationsquellen werden informelle Mittel in hohem Maße sowohl von Bürgern als auch von Einsatzkräften genutzt – insbesondere wenn zentrale Informationen fehlen –, um aktuelle Informationen zu sammeln und auszutauschen, um den Überblick zu behalten, Hilfe zu organisieren, mit Freunden und Familie in Kontakt zu bleiben und aktive oder moralische Unterstützung anzubieten [5]. Neben sozialen Medien, die seit ihrem Aufkommen verstärkt in Krisen genutzt und untersucht werden, unterstützen Multikanalsysteme die Verbreitung von Notfallwarnungen, z.B. via SMS, E-Mail und RSS [13]. Eine aktuelle Studie ergab indes, dass 16% der Bürger in Deutschland Krisen-Apps benutzen [14]. Mit einer wachsenden Anzahl an Medienkanälen und einer schnellen Verbreitung von Inhalten besteht jedoch die Gefahr einer Informationsüberflutung und Fake News, unklarer Quellen und unvollständiger Informationen [5], [15]. Die Benutzerfreundlichkeit und Glaubwürdigkeit eines Kommunikationskanals haben einen großen Einfluss auf dessen Nutzung [5]: Betroffene Parteien nutzten primär traditionelle Medien wie TV, Radio und Zeitungen, da diese im Allgemeinen als glaubwürdigste Quellen gelten [5], [16]. Auch direkte Kommunikation durch persönliche Gespräche, Telefonate und SMS werden als wichtig empfunden [5].

Um bei einem Infrastrukturausfall die gewünschte Reaktion einzuleiten, müssen KRITIS-Betreiber die Menschen über aktuelle Entwicklungen auf dem Laufenden halten, einen Dialog herstellen und während der Krise ehrlich, ansprechbar und glaubwürdig erscheinen. Zudem sollten ihre Mitteilungen mit vertrauenswürdigen Referenzen übereinstimmen [12]. Menge, Art, Medium und Quelle von Informationen haben einen großen Einfluss auf die Wahrnehmung der Krise, der dahinterstehenden Organisation sowie auf die Resonanz der Bevölkerung [5], [15]. Während einer Krise sollten Aktualisierungen bzw. Statusmeldungen, vor allem über die Dauer des Ausfalls, möglichst schnell erfolgen, um emotionale Strapazen, negative Reaktionen, Missverständnisse und unvorhersehbares Verhalten zu vermeiden – vor allem bei Widersprüchen zwischen den Quellen [17]. Eine Umfrage in vier europäischen Ländern zeigt, dass je mehr Teilnehmer soziale Medien nutzen, desto höher die Erwartung dieser Personen ist, dass KRITIS-Betreiber solche Plattformen verwenden, um Informationen bezüglich des anhaltenden Infrastrukturzusammenbruchs zu teilen [16]. In der Studie konnte ein signifikanter Altersunterschied in Bezug auf die Nutzung sozialer, nicht aber traditioneller Medien festgestellt werden.

Forschungsdesiderate. Wenn KRITIS ausfallen, konzentriert sich die Forschung vor allem auf die komplexen Auswirkungen und die Bewältigung des Ausfalls, aber auch auf die Kontinuität und Widerstandsfähigkeit der betroffenen Bevölkerung. Während es Studien darüber gibt, auf welche Art Notfallinstitutionen wie z. B. Polizei und Feuerwehr kommunizieren und auf Infrastrukturzusammenbrüche reagieren, wurde die zivile Kommunikations- und Kontinuitätsaufrechterhaltung – insbesondere in Mitteleuropa – nicht umfassend untersucht [16]. So ist beispielsweise nicht hinreichend erforscht, inwieweit die Öffentlichkeit in diesen Bereichen über ihre lokale Infrastruktur und Krisenmaßnahmen – insbesondere alternative Infrastrukturen – informiert ist. Auch welche Art von Informationen die Öffentlichkeit vom KRITIS-Betreiber erwartet und über welche Kanäle diese kommuniziert werden, erfordert eine genauere Betrachtung. Zudem liegen viele relevante Studien bereits über ein Jahrzehnt zurück, weshalb Bedarf an einer Betrachtung im Kontext aktueller technischer Entwicklungen und entsprechender Einstellungen der Nutzer besteht. Eine Studie, die ähnlichen Fragen nachging, bildet allerdings die europäische Bevölkerung nicht demografisch ab [16]; dies wollen wir durch eine repräsentative Studie ergänzen, indem wir ein möglichst breites Spektrum an Teilnehmern verschiedener Alters- und Einkommensgruppen, Bildungsgrade und verschiedener Herkunft (Bundesländer) befragt haben.

3 Methodik

Die Daten dieses Beitrags stammen aus einer repräsentativen Onlinebefragung, die wir im Juli 2017 in Deutschland unter Nutzung des ISO-zertifizierten Panelproviders GapFish (Berlin) durchgeführt haben. Zur Beantwortung der Forschungsfragen verwenden wir drei geschlossene und eine offene Frage (siehe Anhang), die sich bzgl. der Fragen zu Stromausfällen bei [18] sowie bei Fragen zu Erwartungen Infrastrukturanbietern gegenüber bei [16] orientiert haben, und sich auf das Verhalten und die Mediennutzung während eines Infrastrukturausfalls beziehen. Die Teilnehmer wurden zur Nutzung verschiedener Medienkanäle in Krisensituationen befragt (Q1) bzw. welche Art der Kommunikation sie von KRITIS-Betreibern bei einem Ausfall erwarten (Q2). Überdies sollten sie ihre Kenntnisse in Bezug auf die von Behörden ergriffenen Maßnahmen im Falle eines Notrufausfalls angeben (Q3). Die letzte, offene Frage bot Raum zur Beschreibung des eigenen Verhaltens im Notfall (Q4). Da Krisen durch Infrastrukturausfälle selten sind, konnten die Fragen auch hypothetisch beantwortet werden – unabhängig davon, ob eine solche Situation schon erlebt wurde oder nicht.

Studienteilnehmer. Die Stichprobe der Befragten (N=1.024) wurde an die Verteilung von Alter, Region, Bildung und Einkommen entsprechend der allgemeinen deutschen Bevölkerung [19–21] angepasst. Entsprechend dieser Statistik bestand unsere Stichprobe aus 49,5 % weiblichen und 50,5% männlichen Befragten zwischen 18 und 64 Jahren, von denen fast die Hälfte 45 Jahre und älter war (48%). Wir befragten Teilnehmer aus allen Bundesländern, wobei die größten Teilnehmerzahlen aus Nordrhein-Westfalen (22%) und Bayern (16%) stammten. Nur 1% der Teilnehmer hatte keinen Schulabschluss; 15% besaßen einen Hochschulabschluss. Die Mehrheit verdiente monatlich zwischen 1.500€ und 3.500€ brutto. Ebenso erhoben wir Gewohnheiten zur Techniknutzung mittels einer Fünf-Punkte-Skala (von „stündlich" bis „niemals"). Für das Smartphone gaben fast die Hälfte aller Teilnehmer an, dieses täglich zu nutzen (49%). Ein ähnliches Ergebnis fanden wir bezüglich der täglichen Nutzung von sozialen Medien, z.B. Facebook (46%), Messengern (43%) und YouTube (28%). Insgesamt 33% gaben sogar eine stündliche Nutzung von Smartphones und Messengern an.

Datenanalyse. Für unsere Analyse haben wir zunächst unvollständige Datensätze eliminiert und die Teilnehmerzahl von N=1.069 auf N=1.024 reduziert. Zudem haben wir demografische Variablen wie Alter und Einkommen zu Kategorien zusammengefasst, um einen besseren Vergleich zu ermöglichen. Im Anschluss haben wir die Häufigkeit und Prozentsätze der Antworten auf die geschlossenen Studienfragen in Microsoft Excel berechnet. Zur statistischen Analyse der Daten haben wir das Softwarepaket IBM SPSS Statistics 25 verwendet. Nichtparametrische Tests wurden auf Grundlage von Ordinaldaten ausgewählt. Die Chi-Quadrat-Tests dienten dazu, signifikante Unterschiede zwischen den demografischen Merkmalen sowie den Gewohnheiten und Einstellungen zur Mediennutzung zu untersuchen. Korrelationen zwischen Variablen wurden unter Benutzung des Spearman-Rangkorrelationskoeffizienten bestimmt. Für die qualitative Analyse der offenen Frage wurde die Methode des Open Coding [22] angewandt. Die daraus resultierenden Erkenntnisse stellen eine wichtige Ergänzung der Ergebnisse aus der quantitativen Analyse dar. Nachfolgend werden die Befunde unserer Analysen entsprechend der Forschungsfragen präsentiert.

4 Ergebnisse

4.1 Mediennutzung während Krisen

Eine unserer Fragen bezog sich darauf, welche Informationsquellen die Befragten in einer von ihnen erlebten Krisensituation genutzt haben bzw. welche für sie besonders nützlich waren (siehe Tabelle 1). Für ein Drittel der Teilnehmer (34%) war das Fernsehen eine sehr hilfreiche Quelle während einer Krise, gefolgt von Radio (30%), sozialen Medien (20%), persönlicher Kommunikation (17%) und Notrufdiensten (15%). In der Summe nutzen die meisten Personen den Fernseher (82%) und führten persönliche Gespräche mit Freunden, Familie und Nachbarn (74%). Weitere 74% hörten Radio und 61% führten Telefonate. An fünfter Stelle standen soziale Medien, die von 55% der Befragten genutzt wurden, dicht gefolgt von Zeitungen und Zeitschriften (54%). Andere Medien wurden von weniger als der Hälfte der Befragten verwendet. Krisen-Apps waren der am wenigsten genutzte Medienkanal (25%).

Tabelle 1. Ergebnisse zu Q1: Welche der folgenden Möglichkeiten haben Sie in einer akuten Krisensituation (als Sie selbst betroffen waren / als Sie Helfer(in) waren) genutzt, um sich über die Ereignisse zu informieren, und welche waren für Sie sehr hilfreich?

	Ja, war sehr hilfreich	*Ja, habe ich genutzt*	*Nein*
Fernsehen	34%	48%	18%
Persönliche Gespräche	17%	57%	26%
Radio	30%	44%	27%
Telefonische Gespräche	14%	47%	38%
Soziale Medien	20%	35%	46%
Zeitungen/Zeitschriften	12%	42%	46%
Andere Internetangebote	12%	31%	58%
Kontakt zu Hilfsdiensten	15%	22%	63%
Informationen vor Ort	9%	25%	66%
Krisen-App	10%	15%	74%

Alle Quellen korrelierten signifikant miteinander ($p<.01$). Auch hinsichtlich demografischer und soziotechnischer Faktoren konnten wir signifikante Einflüsse auf die Antworten finden. Dazu fassten wir die Antwortmöglichkeiten in drei Gruppen zusammen: traditionelle Medien (Fernsehen, Radio und Zeitung/Zeitschriften), digitale Medien (soziale Medien, Krisen-Apps und Webseiten) und persönliche Kommunikation vor Ort (Informationen vor Ort, persönliche Gespräche und Hilfsdienste, siehe Tabelle 2). Geschlecht und Region hatten keinen signifikanten Einfluss auf die Nutzung der Informationskanäle, Alter und Bildung dagegen schon: Insbesondere ältere Teilnehmer nutzen das Radio in deutlich höherem Maße ($r=.104$), zeigten aber gleichzeitig eine geringere Tendenz zur Nutzung von sozialen Medien ($r=-.335$), Webseiten ($r=-.205$) oder persönlichen Gesprächen mit Freunden und Familie ($r=-.181$). Die Faktoren junges Alter ($r=.157$), höherer Bildungsabschluss ($r=.053$) und niedriges Einkommen ($r=-.0.58$) führten zu einer größeren Tendenz, sich auf lokale Quellen zu verlassen. Weiterhin hatte die Nutzung von Smartphones einen positiven Effekt auf die Nutzung von Informationen vor Ort (r=.129) und digitalen Medien (r=.296). Häufige Nutzer sozialer Medien nahmen lokale (r=.283) und traditionelle Quellen (r=.115) häufiger in Anspruch. Die Häufigkeit des Postings beeinflusste sowohl die Nutzung von traditionellen (r=.152), digitalen (r=.365) als auch örtlichen Medien (r=.260).

Tabelle 2. Signifikante Ergebnisse der Chi-Quadrat-Tests zur Mediennutzung in Krisen (*p<.05, **p<.01, N=1024)

	Trad. Medien	***Digitale Medien***	***Vor-Ort-Komm.***
Alter		$\chi^2(30,N)=121.64^{**}$	$\chi^2(40,N)=70.32^{**}$
Bildung		$\chi^2(30,N)=53.16^{**}$	$\chi^2(40,N)=64.81^{**}$
Einkommen (EK)			$\chi^2(24,N)=37.62^{*}$
Smartphone-Nutzung (SP)		$\chi^2(24,N)=111.75^{**}$	$\chi^2(32,N)=49.00^{*}$
Social-Media-Nutzung (SM)	$\chi^2(180,N)=274.15^{**}$		$\chi^2(240,N)=424.12^{**}$
Posting-Verhalten (PV)	$\chi^2(24,N)=63.69^{**}$	$\chi^2(24,N)=171.76^{**}$	$\chi^2(32,N)=126.68^{**}$

4.2 Erwartungen an KRITIS-Betreiber in Krisen

Zudem wollten wir herausfinden, welche Informationen die Teilnehmer im Falle eines Infrastrukturausfalls von KRITIS-Betreibern erwarten (siehe Tabelle 3). Die meisten Befragten nannten Mitteilungen über traditionelle Medien (72%) und die Webseite des KRITIS-Betreibers (70%). Darüber hinaus erwartete die Mehrheit eine SMS (60%), eine Nachricht über soziale Medien (59%), einen Anruf (58%) oder eine persönliche Ansprache vor Ort (57%). Zwei-Wege-Kommunikation erachteten die meisten (60%) als neutral oder unnötig. Die zentralsten Erwartungen an den Inhalt der Informationen umfassten Instruktionen (83%), die Ursache des KRITIS-Ausfalls (80%) sowie die voraussichtliche Dauer der Störung (78%). Für 70% war die Quelle der Informationen von Relevanz.

Alle Antworten korrelieren signifikant sowohl untereinander als auch mit den vorherigen Fragen (p<.0001). Die Nutzung von sozialen Medien während Krisen (r=.388, p<.0001) beeinflusste die Erwartungen einer Mitteilung auf den sozialen Kanälen des KRITIS-Betreibers maßgeblich. Das gleiche Ergebnis fand sich auch in Bezug auf traditionelle Medien, z.B. Zeitungen (r=.070, p<.05), Fernsehen (r=.105, p<.0001) und Radio (r=.159, p<.002). Zudem konnten wir feststellen, dass ein jüngeres Alter die Erwartungen eines Telefonanrufs (r=-.098) und einer Nachricht in sozialen Medien (r=-.114) erhöht, jedoch das Gegenteil für eine Meldung in traditionellen Medien gilt (r=.160). Je älter die Teilnehmer waren, desto mehr erwarteten sie, dass der KRITIS-Betreiber Anweisungen veröffentlicht (r=.088), die Quelle der Information angibt (r=.089) und Informationen über die Dauer des Ausfalls zur Verfügung stellt (r=.102). Zudem fanden wir auch Einflüsse des Einkommens und Bildungsstatus: Speziell auf Mitteilungen über traditionelle Medien (r=.099 und r=.085) und Webseiten (r=.116 und r=0.94) gab es einen signifikant positiven Einfluss. Ein höheres Einkommen erhöhte auch die Informationserwartung, insbesondere über die Dauer des Ausfalls (r=.115). Alle signifikanten Ergebnisse der Chi-Quadrat-Tests sind in Tabelle 4 zu finden.

Tabelle 3. Ergebnisse zu Q2: Was erwarten Sie bei Ausfällen der Infrastruktur (z.B. Strom, Gas oder Telekommunikation) seitens des Betreibers (z.B. Telekom, Energieversorger)?

	Starke Zust.	***Zust.***	***Neutral***	***Abl.***	***Starke Abl.***
Instruktionen	45%	38%	15%	2%	1%
Informationen zum Grund	45%	35%	17%	3%	1%
Informationen zur Dauer	45%	33%	17%	3%	1%
Meldung in trad. Medien	35%	37%	21%	5%	2%
Meldung auf Website	35%	35%	21%	6%	3%
Quelle der Information	32%	38%	26%	3%	1%
SMS-Nachricht	25%	35%	29%	8%	4%
Social Media Post	25%	34%	25%	9%	6%
Anruf	22%	36%	31%	7%	3%
Ansprache vor Ort	22%	35%	33%	8%	2%
Zwei-Wege-Komm.	13%	28%	51%	6%	3%

Tabelle 4. Signifikante Chi-Quadrat-Ergebnisse zur Kommunikationserwartung in Krisen (*p<.05, **p<.01, N=1024), siehe Tabelle 2 zur Erläuterung von Abkürzungen

	Anruf	***SMS-Nachricht***	***Social Media Post***	***Website***
Alter	$\chi^2(20,N)=43.13^{**}$			
Bundesland		$\chi^2(56,N)=77.95^{*}$		
Bild.	$\chi^2(20,N)=41.95^{**}$			
EK				$\chi^2(12,N)=24.96^{*}$
SP		$\chi^2(16,N)=35.01^{**}$	$\chi^2(16,N)=44.65^{**}$	
SM	$\chi^2(120,N)=175.96^{**}$		$\chi^2(1080,N)=1407.23^{**}$	
PV	$\chi^2(16,N)=31.15^{*}$	$\chi^2(16,N)=28.54^{*}$	$\chi^2(16,N)=144.91^{**}$	$\chi^2(16,N)=35.18^{**}$

	Vor Ort	***Trad. Medien***	***Zwei-Wege***	***Dauer***	***Grund***
Alter		$\chi^2(20,N)=52.73^{**}$		$\chi^2(20,N)=41.20^{**}$	
Bild.	$\chi^2(20,N)=43.07^{**}$	$\chi^2(20,N)=46.52^{**}$			
EK		$\chi^2(12,N)=21.48^{*}$			
SM			$\chi^2(120,N)=159.77^{**}$		
PV	$\chi^2(16,N)=34.92^{**}$		$\chi^2(16,N)=32.63^{**}$		$\chi^2(16,N)=30.48^{*}$

4.3 Maßnahmen während eines Ausfalls der Notrufnummer

Überdies wollten wir herausfinden, über welches Wissen die Teilnehmer in Bezug auf Maßnahmen während eines KRITIS-Ausfalls verfügen. Die Mehrheit wusste, wo sich die nächste Polizeistation (70%) und andere Verwaltungsgebäude befinden (68%), kannte den Standort der nächstgelegenen Rettungswache (55%) und des Feuerwehrgerätehauses (55%). Mehr als die Hälfte war über die Tatsache, dass Medien über alternative Maßnahmen berichten (56%) und die Polizei ihre Aktivität im betroffenen Gebiet erhöht (51%), informiert. Hingegen war der Standort des nächstliegenden Gebäudes des Technischen Hilfswerks (THW) nur 27% der Befragten bekannt, und knapp ein Viertel (23%) der Teilnehmer wusste, dass ein Fahrzeug der Feuerwehr an Orten ohne

Feuerwehrwache auf einem zentralen Platz anzufinden ist. Nur 19% waren sich der Tatsache bewusst, dass im Krisenfall die Feuerwachen durch BOS besetzt sind.

Tabelle 5. Ergebnisse zu Q3: Bei einem Ausfall der Notrufnummer 112 werden verschiedene Maßnahmen getroffen, um der Bevölkerung weiterhin die Möglichkeit zu geben, einen Notfall zu melden. Zu welchem Grad sind Ihnen die folgenden Maßnahmen bzw. Standorte bekannt?

	Sehr bek.	*Bek.*	*Neutral*	*Unbek.*	*Sehr unbek.*
Standort Polizeiwache	29%	41%	14%	10%	7%
Standort Verwaltung	31%	37%	16%	10%	7%
Alternativen in Medien	13%	43%	22%	14%	8%
Standort FW-Gerätehaus	24%	31%	18%	16%	10%
Standort Rettungswache	22%	33%	21%	15%	10%
Verstärkte Polizeipräsenz	13%	38%	20%	18%	12%
Standort THW	10%	17%	21%	34%	17%
Orte ohne FF oder BF mit FW-Fahrzeugen besetzt	6%	17%	17%	35%	25%
FW-Gerätehäuser besetzt	7%	12%	15%	36%	30%

Die Antworten für alle Maßnahmen korrelieren signifikant miteinander und mit der Nutzung aller Kommunikationskanäle in Krisen ($p<.0001$). Das Verlassen auf Telefongespräche, persönliche Kontakte und Hilfsdienste zeigte die höchste Korrelation ($r=.289$, $r=.282$ und $r=.273$). Zur Untersuchung des möglichen Einflusses von demografischen Faktoren sowie der Nutzung sozialer Medien auf die Antworten wurden diese zu einem einzigen Wert zusammengefasst und darauf basierend Chi-Quadrat-Tests durchgeführt. Hier beobachteten wir einen signifikanten Einfluss von Bildung ($\chi^2(180, 1024)=225.51$; $p<.05$; $r=.006$) und Einkommen ($\chi^2(108, 1024)=1407.23$; $p<.0001$; $r=.047$), obwohl ein Trend aufgrund schwacher Spearman-Korrelationen nicht eindeutig festzustellen ist. Die Nutzung von Smartphones ($\chi^2(144, 1024)=178.25$; $p<.05$; $r=.037$) und sozialen Medien ($\chi^2(1080, 1024)=225.51$; $p<.05$; $r=.202$) als auch die Häufigkeit von Postings in sozialen Netzwerken ($\chi^2(144, 1024)=224.53$; $p<.0001$; $r=.141$) wiesen dagegen eine signifikante Korrelation mit dem Wissen auf.

Etwa ein Drittel der Teilnehmer (n=344, 33,59%) hatten im Durchschnitt eine geringe Kenntnis der Maßnahmen (neutral oder schlechter). Im Vergleich zur Gesamtstichprobe sind die demografischen Werte dieser Gruppe sehr ähnlich verteilt, ein t-Test für unabhängige Stichproben zeigte dabei einige signifikante Unterschiede: Diese Gruppe war etwas älter ($t(1022)=2.47$, $p=.014$), benutzte soziale Medien seltener ($t(1022)=5.09$, $p<.000$) und postete weniger häufig darin ($t(1022)=2.39$, $p=.017$).

4.4 Antizipiertes Verhalten in Krisensituationen

Um Lösungen für eine bessere IKT-Unterstützung und effektive Kommunikationsstrategien für KRITIS-Betreiber im Falle von Infrastrukturausfällen zu erarbeiten, haben wir die Teilnehmer gebeten, ihr Verhalten in Krisensituationen detailliert in einer offenen Frage zu beschreiben (siehe Anhang). Anhand des Open Coding [22] haben wir Antworten einem oder mehreren Codes zugewiesen, die die Aussage beschreiben und iterativ verfeinert wurden. Am Ende der Kodierung zeichneten sich 41 individuelle Codes ab, die wir in sieben Kategorien, im Folgenden kursiv hervorgehoben, zusammengefasst haben. Die Gesamtzahlen der Aussagen in den einzelnen Kategorien sind wie folgt: Kommunikation: 692; Information: 646; Persönliche Sicherheit: 490; Keine Antwort: 209; Materielle Sicherheit: 145; Probleme: 140; Passives Verhalten: 112.Im Hinblick auf *persönliche Sicherheit* gab etwa ein Fünftel (n=178) an, dass sie Schutz suchen oder zu Hause bleiben würden, sofern dies sicher sei. 10% der Befragten nannten auch Familienmitglieder, um die sie sich kümmern würden. Darüber hinaus würden 158 Teilnehmer auch anderen außerhalb der eigenen Familie, wie etwa Nachbarn, Hilfe anbieten und beispielsweise nach gefährdeten Menschen Ausschau halten. Auch die Warnung anderer – ob persönlich, per Anruf, Nachricht oder in sozialen Netzwerken – war eine weitverbreitete Reaktion. In Bezug auf die *materielle Sicherheit* spielte die Versorgung mit Lebensmitteln und anderen Ressourcen für 78 Individuen eine wichtige Rolle. Insgesamt 13 der Befragten waren sich der Notwendigkeit der Vorbereitung bewusst. Der zweite Aspekt dieser Kategorie setzte sich aus persönlichen Gegenständen zusammen, z.B. Dokumenten und Schlüsseln (n=67).

In der Kategorie *Kommunikation* stand die Warnung Angehöriger und Weiterleitung von Informationen im Vordergrund. Darauf folgte die Kontaktaufnahme mit anderen Menschen im Allgemeinen (n=63). Familie (n=176), Freunde (n=81) und Nachbarn (n=65) waren hierbei häufig genannte Adressaten, um vor allem den Sicherheitsstatus auszutauschen. Nach Rat zu fragen, Zusammenarbeit anzubieten und Informationen auszutauschen (n=47) waren weitere Gründe zur erweiterten Kontaktaufnahme. In der Kategorie *Information* fanden wir heraus, dass mehr als die Hälfte der Teilnehmer (n=447) Informationen über die jeweilige Krise sammeln würden. Von denjenigen Befragten, die einen bestimmten Medienkanal angaben, nannten 189 Personen traditionelle, während etwas weniger die Nutzung digitaler Medien erwähnten (n=136). Zusätzlich gaben 69 Teilnehmer an, dass sie Passanten oder Helfer vor Ort befragen oder Lautsprecherdurchsagen Gehör schenken würden. Andere würden sich direkt an Notfallorganisationen, Regierungsvertreter oder KRITIS-Betreiber wenden (n=44). Währenddessen erklärten 83 Teilnehmer explizit, dass sie auf durch die Medien und Hilfsdienste kommunizierte Anweisungen warten und diese befolgen würden.

In den letzten drei Kategorien fassten wir Codes zusammen, welche keine Handlungen enthielten, die die Teilnehmer im Falle einer Krise unternehmen würden. Leere Antworten und Aussagen wie „Keine Ahnung“ oder „Weiß ich nicht“ wurden in eine eigene Kategorie *keine Antwort* aufgenommen. Unter *passives Verhalten* fiel etwa nichts tun (n=32) und Ruhe bewahren (n=77). Darüber hinaus stießen wir auf *Probleme*, welche die Teilnehmer bei der Beantwortung der offenen Frage hatten. Die Teilnehmer wiesen darauf hin, dass ihr Handeln von der Art der Krise abhängen würde (n=106). 22 Befragte gaben ausdrücklich an, noch nie zuvor eine Krise erlebt zu haben und deshalb keine qualifizierten Antworten geben zu können. Ebenso verwiesen neun Menschen darauf, dass sie in einem tatsächlichen Krisenfall anders reagieren würden und sie ihre Reaktionen deshalb nicht vorhersagen könnten.

5 Diskussion und Fazit

Der Ausfall kritischer Infrastrukturen beeinträchtigt viele Ressourcen- und Kommunikationskanäle, behindert den Transfer notwendiger Informationen und verursacht wirtschaftlichen Schaden [23]. Obwohl sich die Forschung mit den Eigenschaften, der Resilienz sowie den kooperativen und technischen Aspekten von KRITIS [15] beschäftigt hat, wurde die Einstellung von Bürgern gegenüber KRITIS-Ausfällen und IKT-Nutzung zu deren Ausgleich in Mitteleuropa noch nicht repräsentativ untersucht. Unsere Studie trägt daher mit spezifischen Aspekten des Bewusstseins, Verhaltens und der Kommunikation der deutschen Bevölkerung in KRITIS-Ausfällen dazu bei. Auch wenn unsere Stichprobe nur hinsichtlich demografischer Aspekte repräsentativ ist, ergeben sich aus unseren Ergebnissen in einem wenig beachteten Feld aktuelle und unter verschiedenen Gesichtspunkten untersuchte Aussagen, die helfen können, ein Gesamtbild zum Nutzen von Medien in KRITIS-Ausfällen in Deutschland zu zeichnen und weitere Forschung anzuleiten.

Bewusstsein und Verhalten in KRITIS-Ausfällen (FF1). Frühere Studien zeigten bereits, dass die Mehrheit der Bürger noch keinen schwerwiegenden KRITIS-Ausfall miterlebt hat [3], [24]. Etwa 10% unserer Teilnehmer hatten keine Antwort auf die Frage, was sie in einem Notfall tun würden, während andere glaubten, dass die Maßnahmen abhängig von der Art der Krise und damit nicht vorherzusagen seien. Dennoch können wir ein steigendes Gefahrenbewusstsein beobachten, vor allem bedingt durch Terrorattacken. Auf die Frage zu ihrem Krisenverhalten gaben die meisten Teilnehmer an, auf bekannte Medienkanäle zurückgreifen, um Informationslücken zu füllen und Handlungsempfehlungen zu bekommen. Im Fall einer Krise haben die eigene Sicherheit und die der Angehörigen, der Austausch von Informationen und die gegenseitige Hilfe oberste Priorität. Da die effektive Vorbereitung und Bewältigung von Krisen in Deutschland in den Händen der Behörden und Einsatzkräfte liegt, ist deren Aufgabe vor allem, die Bevölkerung für potenzielle Gefahren zu sensibilisieren, anzuleiten und einen großen Schaden durch mangelnde Vorbereitung und Risikobewusstsein zu vermeiden [3]. Hingegen kennen viele unserer Teilnehmer Gebäude und Maßnahmen, die sie im Fall eines KRITIS-Ausfalls nutzen können. Dies korreliert sogar mit der Nutzung sozialer Medien und anderem Verhalten wie dem Kontaktieren von Hilfsdiensten und der Kommunikation mit anderen per Telefon. Dies zeigt, dass diese Medien effektiv zur Krisenvorbereitung genutzt werden könnten.

Kommunikation und Information durch Medienkanäle (FF2). Im Vergleich zu sozialen Medien und Webseiten waren traditionelle Medien wie Fernsehen und Radio die am häufigsten genannten Informationsquellen [5]. Dies ist vermutlich darauf zurückzuführen, dass sie als offiziell, aktuell und vertrauenswürdig wahrgenommen werden, was einigen Teilnehmern wichtig war. In Übereinstimmung mit einer qualitativen Studie [15] tendierten viele Befragte dazu, Augenzeugen vor Ort zu vertrauen, und persönliche Kommunikation wurde von über 70% aller Teilnehmer praktiziert. Dennoch wird letztere selten als sehr hilfreiche Quelle genannt – im Gegensatz zu sozialen Medien. Im Vergleich zu einer Studie aus 2017 nahmen die Probanden soziale Medien in dieser Befragung sogar häufiger in Anspruch (55%), während die Nutzung von Krisen-Apps auf von 16% auf 25% anstieg [9]. Obwohl diese noch nicht weit verbreitet sind, nahmen 40% ihrer Nutzer sie als sehr hilfreich wahr. KRITIS-Betreiber könnten in solchen Apps daher über Institutionen, Maßnahmen und mögliche Risiken informieren als auch darauf hinweisen, wie man sich vorbereiten und verhalten sollte. Einflüsse der Demographie und IKT-Nutzung in Krisensituationen, die in bisherigen Arbeiten gefunden wurden [16], konnten in dieser repräsentativen Stichprobe für diese und die folgende Frage repliziert werden. Jedoch war das nicht der stärkste Einfluss: Je häufiger Bürger bestimmte IKT im Alltag und in Krisen nutzten, desto stärker sahen sie ihren Mehrwert in Krisen. IKT dienen vor allem dazu, Informationen zu sammeln, Familie und Freunde zu kontaktieren sowie Informationen und Sicherheitsstatus auszutauschen.

Erwartungen an KRITIS-Betreiber (FF3). Bezüglich der Kommunikation mit KRITIS-Betreibern schätzten die Befragten fast jede Art von Medien und über die Hälfte der Teilnehmer erwartete, dass KRITIS-Betreiber soziale Medien in Krisensituationen nutzen. Die Bereitstellung von Informationen, insb. Handlungsempfehlungen, sowie Dauer und Ursachen der Krise, wurde als wichtiger bewertet als die Etablierung einer Zwei-Wege-Kommunikation [9].

Implikationen für Design und Einsatz von IKT (FF4). Unsere Ergebnisse zeigen, dass es ein starkes Bedürfnis nach Informationen über und zur Sensibilisierung für Krisen gibt. Traditionelle und soziale Medien werden als hilfreiche und zuverlässige Kanäle für die Verbreitung dieser Informationen angesehen. Jedoch erlaubt in bestimmten Infrastrukturausfällen (z. B. Strom) die Akkulaufzeit von Smartphones weiterhin eine begrenzte Nutzung von Krisen-Apps, welche mitunter Checklisten und Verhaltensinformationen vor, während und nach verschiedenen Arten von Notfällen beinhalten [14]. Daher kann eine Verbreitung geeigneter IKT die soziale Widerstandsfähigkeit und Bewusstsein für Verhaltensmaßnahmen erhöhen [8]. Zudem könnte dies auch das generelle Verständnis von Infrastrukturen als wichtige Basis für viele Bereiche der Kooperation erhöhen [11], [23]. Da demografische Eigenschaften und Techniknutzung einen Einfluss auf die Antworten haben, sollten Kommunikationskanäle und Informationen an ihre jeweilige Zielgruppe angepasst werden [10]. Gleichzeitig sollte verstärkt auf schnelle, zugängliche, korrekte, vollständige und belegte Information geachtet werden, um mithilfe von IKT eine effektivere Krisenbewältigung und Kooperation zwischen KRITIS-Betreibern, Behörden und Bürgern aufzubauen.

Abschließende Betrachtungen. Zu einem hohen Grad unterstützen unsere Ergebnisse die Aussagen bisheriger Studien. Im Vergleich zu bestehenden Untersuchungen liefern wir jedoch aktuelle Einblicke in den wenig erforschten Bereich der Einstellung, Erwartung und Nutzung bezüglich Medien in Ausfällen stabiler KRITIS anhand einer demografisch repräsentativen Stichprobe aus ganz Deutschland. Unter Einbezug diverser Faktoren wie der alltäglichen Techniknutzung bietet unsere Forschung eine Grundlage für Vergleiche unter anderem mit weiteren Arten von Zielgruppen, Krisen, Einflussbereichen und Ländern. Bei einem Ausfall bestimmter Kanäle kann die Relevanz anderer zunehmen bzw. diese vorübergehend ersetzen. Dabei gilt es, möglichst viele Kanäle abzudecken und auch die Instruktionen je nach Art der Krise anzupassen. Dabei wurde deutlich, dass IKT für Krisenkommunikation in folgenden Bereichen sinnvoll eingesetzt werden kann:

- Information über potentielle Gefährdungen (Risikokommunikation) und Verbesserung der Resilienz der Zivilbevölkerung (Handlungsempfehlungen für persönliche und materielle Sicherheit, Notbevorratung, etc.).
- Während eines KRITIS-Ausfalls: Kommunikationsunterstützung zwischen Akteuren (BOS, KRITIS-Betreiber, Zivilbevölkerung) auf multiplen Kanälen.
- Während und nach einem Ausfall der Infrastruktur: Bereitstellung aktueller, lokaler, spezifischer, vertrauenswürdiger und zielgruppenorientierter Informationen durch BOS und KRITIS-Betreiber.

Die Studie unterliegt einigen Limitationen: Da Krisen wie Infrastrukturausfälle in Deutschland selten sind, waren unsere Fragen hypothetisch gestellt, sodass empirisch geprüft werden muss, inwiefern sie mit tatsächlichem Krisenverhalten übereinstimmen. Bei bestehenden Studien zu tatsächlichen Infrastrukturausfällen wurde etwa bereits mangelnde Vorbereitung seitens der Bürger und erhöhter Informationsbedarf festgestellt [25]. Eine Antwortoption für Teilnehmer, die noch nicht von Krisen betroffen waren, sollte ergänzt werden, um genauere und valide Ergebnisse zu erzielen. Ein Fokus auf bestimmte Ereignisse, in diesem Fall KRITIS-Ausfälle, ist außerdem notwendig, um Verhalten und Erwartungen in verschiedenen Situationen abzugrenzen. Weiterhin würden Begriffserklärungen und klarere Abgrenzung von Skalen zum besseren Verständnis der Fragebögen beitragen. Die Tatsache, dass wir unsere Probanden online befragt haben, schränkt zudem die Repräsentativität ein. Daher müsste die Befragung offline und mit Teilnehmern unterschiedlicher Technikaffinität durchgeführt werden. Darüber hinaus könnte man die Gründe hinter dem Verhalten und der Mediennutzung, die in der offenen Frage angedeutet werden, genauer untersuchen, um die Intentionen und Bedürfnisse der Betroffenen besser zu verstehen.

Danksagung: Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen von KontiKat [26] (13N14351), teilweise im Rahmen von HyServ (01IS17030B) sowie durch die DFG innerhalb des SFB 1053 MAKI finanziert. Zudem hat Marius Brüser an der Konzeption des Fragebogens mitgewirkt.

Literaturverzeichnis

1. Guha-Sapir, D., Hargitt, D., Hoyois, P.: Thirty Years of Natural Disasters 1974-2003: The Numbers. Presses universitaires de Louvain, Louvain-la-Neuve (2004)
2. Bundesnetzagentur und Bundeskartellamt: Monitoringbericht 2016. Bundesnetzagentur und Bundeskartellamt, Bonn (2016). https://www.bundesnetzagentur.de/
3. Holenstein, M., Küng, L.: Stromausfall - was denkt die Bevölkerung. Sicherheit. 3, 61 (2008)
4. Hughes, A.L., Palen, L., Sutton, J., Liu, S.B., Vieweg, S.: “Site-Seeing” in Disaster: An Examination of On-Line Social Convergence. In: Proceedings of the 5th International ISCRAM Conference. pp. 1–10. ISCRAM, Brüssel (2008)
5. Austin, L., Liu, B.F., Jin, Y.: How audiences seek out crisis information: Exploring the social-mediated crisis communication model. J. Appl. Commun. Res. 40, 188–207 (2012)

6. Jochimsen, R.: Theorie der Infrastruktur: Grundlagen der marktwirtschaftlichen Entwicklung. Mohr (Siebeck), Tübingen (1966)
7. Rat der Europäischen Union: Richtlinie 2008/114/EG des Rates vom 8. Dezember 2008 über die Ermittlung und Ausweisung europäischer kritischer Infrastrukturen und die Bewertung der Notwendigkeit, ihren Schutz zu verbessern. Amtsblatt der Eur. Union. 345, 75–82 (2008)
8. Boin, A., McConnell, A.: Preparing for critical infrastructure breakdowns: The limits of crisis management and the need for resilience. J. Contingencies Cris. Manag. 15, 50–59 (2007)
9. Reuter, C., Kaufhold, M.-A., Spielhofer, T., Hahne, A.S.: Social Media in Emergencies: A Representative Study on Citizens' Perception in Germany. In: Proceedings of the ACM on Human Computer Interaction (PACM). pp. 1–19. ACM, New York (2017)
10. O'Sullivan, T.L., Kuziemsky, C.E., Toal-Sullivan, D., Corneil, W.: Unraveling the complexities of disaster management: A framework for critical social infrastructure to promote population health and resilience. Soc. Sci. Med. 93, 238–246 (2013)
11. Semaan, B., Mark, G.: Technology-mediated social arrangements to resolve breakdowns in infrastructure during ongoing disruption. ACM Trans. Comput. Interact. 18, 1–21 (2011)
12. Seeger, M.W.: Best practices in crisis communication: An expert panel process. J. Appl. Commun. Res. 34, 232–244 (2006)
13. Klafft, M.: Diffusion of emergency warnings via multi-channel communication systems an empirical analysis. In: Eleventh International Symposium on Autonomous Decentralized Systems (ISADS). pp. 1–5. IEEE, Piscataway, NJ (2013)
14. Reuter, C., Kaufhold, M.-A., Leopold, I., Knipp, H.: KATWARN, NINA or FEMA? Multi-Method Study on Distribution, Use and Public Views on Crisis Apps. In: Twenty-Fifth European Conference on Information Systems (ECIS). pp. 2187–2201. AISeL, Atlanta, GA (2017)
15. Huang, Y.L., Starbird, K., Orand, M., Stanek, S.A., Pedersen, H.T.: Connected through crisis: emotional proximity and the spread of misinformation online. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15). pp. 969–980. ACM, New York (2015)
16. Petersen, L., Fallou, L., Reilly, P., Serafinelli, E.: Public expectations of social media use by critical infrastructure operators in crisis communication. In: Proceedings of the 14th International ISCRAM Conference. pp. 522–531. ISCRAM, Bruxelles (2017)
17. Reynolds, B., Seeger, M.W.: Crisis and emergency risk communication as an integrative model. J. Health Commun. 10, 43–55 (2005)
18. Reuter, C.: Communication between Power Blackout and Mobile Network Overload. Int. J. Inf. Syst. Cris. Response Manag. 6, 38–53 (2014)
19. Statistisches Bundesamt: Bevölkerung Deutschlands nach Altersgruppen 2016. Statista, Hamburg (2016). https://de.statista.com/
20. Bundeszentrale für politische Bildung (bpb): Datenreport 2016. Ein Sozialbericht für die Bundesrepublik Deutschland. Statistisches Bundesamt, Bonn (2016). https://www.destatis.de/
21. Statistisches Bundesamt: Bildungsstand der Bevölkerung - 2016. Statistisches Bundesamt, Wiesbaden (2016). https://www.destatis.de/
22. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine de Gruyter, New York (1967)
23. Pipek, V., Wulf, V.: Infrastructuring: Towards an Integrated Perspective on the Design and Use of Information Technology. J. Assoc. Inf. Syst. 10, 447–473 (2009)

24. Reuter, C., Backfried, G., Kaufhold, M.-A., Spahr, F.: ISCRAM turns 15: A Trend Analysis of Social Media Papers 2004-2017. In: Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). p. accepted. , Rochester, New York, USA (2018)
25. Nestler, S.: Flächendeckende Kommunikation im Stromausfall durch regionale IKT Krisenszenario : Längerfristiger Stromausfall. In: Burghardt, M., Wimmer, R., Wolff, C., and Womser-Hacker, C. (eds.) Mensch und Computer 2017 – Workshopband. pp. 9–16. Gesellschaft für Informatik e.V., Bonn (2017)
26. Reuter, C., Kaufhold, M.-A., Schorch, M., Gerwinski, J., Soost, C., Hassan, S.S., Rusch, G., Moog, P., Pipek, V., Wulf, V.: Digitalisierung und Zivile Sicherheit: Zivilgesellschaftliche und betriebliche Kontinuität in Katastrophenlagen (KontiKat). In: Hoch, G., Schröteler von Brandt, H., Stein, V., and Schwarz, A. (eds.) Sicherheit (DIAGONAL Jahrgang 38). pp. 207–224. Vandenhoeck & Ruprecht, Göttingen (2017)
27. Geenen, E.M.: Kollektive Krisen. Katastrophe, Terror, Revolution - Gemeinsamkeiten und Unterschiede. In: Clausen, L., Geenen, E.M., and Macamo, E. (eds.) Entsetzliche soziale Prozesse. Theoretische und empirische Annäherungen. pp. 5–23. Lit-Verlag, Münster (2003)

Anhang: Fragen

Q1: Welche der folgenden Möglichkeiten haben Sie in einer akuten Krisensituation (als Sie selbst betroffen waren / als Sie Helfer(in) waren) genutzt, um sich über die Ereignisse zu informieren, und welche waren für Sie sehr hilfreich? (Ja, sehr hilfreich; Ja, habe ich genutzt; Nein): 1. Zeitungen und Zeitschriften | 2. Fernsehen | 3. Radio | 4. Persönliche Gespräche (z.B. mit Familie, Freunden und Nachbarn) | 5. Telefonische Gespräche (z.B. mit Familie, Freunden und Nachbarn) | 6. Kontaktaufnahme zu Rettungsdienst, Feuerwehr, Polizei oder Krankenhaus) | 7. Informationsmöglichkeiten vor Ort (z.B. Aushänge, Flugblätter und Lautsprecherdurchsagen) | 8. Soziale Medien (z.B. Facebook, Twitter, Instagram und YouTube) | 9. Andere Internetangebote | 10. Eine Krisen-App (z.B. KATWARN oder NINA) | 11. Keine | 12. Weiß nicht

Q2: Was erwarten Sie bei Ausfällen der Infrastruktur (z.B. Strom, Gas oder Telekommunikation) seitens des Betreibers (z.B. Telekom, Energieversorger)? (Starke Zustimmung; Zustimmung; Neutral; Ablehnung; Starke Ablehnung): 1. Persönliche Ansprache vor Ort | 2. Anruf | 3. SMS-Nachricht | 4. Nachricht auf sozialen Medien | 5. Meldung auf der Website | 6. Meldung in traditionellen Medien | 7. Zwei-Wege-Kommunikation | 8. Informationen zur Dauer | 9. Informationen zum Grund | 10. Informationen, was ich tun soll | 11. Quelle der Information

Q3: Bei einem Ausfall der Notrufnummer 112 werden verschiedene Maßnahmen getroffen, um der Bevölkerung weiterhin die Möglichkeit zu geben, einen Notfall zu melden. Zu welchem Grad sind Ihnen die folgenden Maßnahmen bzw. Standorte bekannt? (Ist mir sehr bekannt; Ist mir bekannt; Neutral; Ist mir nicht bekannt; Ist mir überhaupt nicht bekannt): 1. Durch die BOS werden Feuerwehrgerätehäuser im betroffenen Gebiet besetzt. | 2. In Ortschaften ohne Feuerwehrgerätehaus wird ein Fahrzeug der Feuerwehr auf einem zentralen Platz positioniert. | 3. Die Polizei verstärkt ihre Aktivität im betroffenen Gebiet. | 4. Durch die Medien werden weitere Alternativen gemeldet. | 5. Standort der nächstliegenden Rettungswache (Standort Krankenwagen etc.) | 6. Standort des nächstliegenden Feuerwehrgerätehauses | 7. Standort des nächstliegenden Gebäudes des Technischen Hilfswerks (THW) | 8. Standort der nächstliegenden Polizeistation | 9. Standort der nächstliegenden Verwaltungsgebäude (z.B. Rathaus, Kreishaus)

Q4: Beschreiben Sie möglichst detailliert, wie Sie in einer außergewöhnlichen Lage (Krise, Katastrophe) vorgehen oder vorgehen würden? Welche Tätigkeiten führen Sie durch oder würden Sie durchführen?

Fake News Perception in Germany: A Representative Study of People's Attitudes and Approaches to Counteract Disinformation

Christian Reuter, Katrin Hartwig, Jan Kirchner, and Noah Schlegel

Technische Universität Darmstadt,
Science and Technology for Peace and Security (PEASEC), Germany
reuter@peaesec.tu-darmstadt.de

Abstract. Fake news has become an important topic in our social and political environment. While research is coming up for the U.S. and European countries, many aspects remain uncovered as long as existing work only marginally investigates people's attitudes towards fake news. In this work, we present the results of a representative study (N=1023) in Germany asking participants about their attitudes towards fake news and approaches to counteract disinformation. More than 80% of the participants agree that fake news poses a threat. 78% see fake news as harming democracy. Even though about half of the respondents (48%) have noticed fake news, most participants stated to have never liked, shared or commented on fake news. Regarding demographic factors, our findings support the view of younger and relatively educated people being more informed about fake news. Concerning ideological motives, the evaluation suggests left-wing or liberal respondents to be more critical of fake news.

Keywords: Fake news, disinformation, citizens' perceptions, representative survey, counteraction, Germany

1 Introduction

Information systems play a crucial role regarding peace and security [1], also with regard to interactive systems [2]. Since the 2016 U.S. presidential election, the term *fake news* is widely known and has found its way into both scientific and public debates. In Germany, the 2017 parliamentary election was accompanied by discussions about fake news. However, research shows that there was no major fake news during the campaign which had any impact on the election results [3]. Although there were many cases of fake news during the U.S. election, scholars argue that they did not have any impact on its outcome [4]. These observations suggest that the perception of fake news influencing people's attitudes diverges from its actual impact. As fake news is apparent online, it may shape users' social media experiences to varying degrees, potentially influencing important social dynamics [5–7]. Thus, we ask for a clearer picture of the population's perception of and interaction with fake news. In the context of this study, fake news is

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

to be understood as "all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit" [8].

Our paper tries to find answers to the following questions: *(1) What **attitudes** do people have towards fake news? (2) Have they ever noticed fake news or **interacted** with it? (3) How do they evaluate possible approaches to **counteract** fake news?* Therefore, we conducted a representative study (N=1023) and analyzed our results referring to demographic factors as well as to perceptions of threats to peace (i.e., ideological standpoints) to ask for influential factors regarding the perception, interaction, and counteraction of fake news. This allowed us to focus on attitudinal and behavioral patterns of different groups. While it is certainly important to examine actual fluctuations of fake news, humans' respective perception may constitute an essential factor for their behavior online or in cases of counteraction. We focus our work on one country permitting a differentiated analysis and presentation of results which may prove relevant with respect to other countries across Europe. Germany is an important European country and just recently had the 2017 parliamentary elections, introducing a right-wing populist party into parliament. Therefore, we performed a study representative of the German adult population under 65 on their opinion on fake news. First, we elucidate the phenomenon of fake news and discuss relevant work (section 2). Next, we introduce the representative survey, describing our approach and methodology (section 3) and presenting the results (section 4). Last, we discuss the results and draw a conclusion (section 5).

2 Background and Related Work

2.1 Fake News as a Social Phenomenon

Before 2016, fake news mainly referred to satirical news shows; the perception changed when a lot of fake news went viral, starting to affect political parties globally and influencing opinions on a larger scale than before [9]. While being such a popular and frequent term, it is often mingled with other phenomena [3]. Its fuzzy meaning facilitates misuses of the term to discredit undesired news [10]. Furthermore, in German debates, its meaning is often mixed with hate speech [3]. Allcott and Gentzkow [4] define fake news as "news articles that are intentionally and verifiably false and could mislead readers". In contrast, misinformation refers to "information that is false but not created with the intention of causing harm" [11]. Fake news often deals with controversial issues like migration, child abuse, or war [12]. Although there is no consensus in science as to whether social media encourages the emergence of "digital echo-chambers" – that is to say: a system that amplifies and reinforces ideas of beliefs at the expense of competing vies – its popularity has been proven to facilitate the spread and success of fake news [13]. Nowadays, many people use social media instead of mainstream media as a source of information [14]. Consequently, the important role of journalism as a gatekeeper has shifted to individuals who must decide on their own whether a report is reliable [15], [16]. Well-studied occurrences of fake news appear in the context of elections although studies show that it had little overall impact on them [3], [4]. Yet, fake news may be used to manipulate public opinion and debate. According to the

German investigative journalism collective *Correctiv*, most fake news in Germany has originated from supporters and politicians of the right-wing populist party *Alternative f r Deutschland* (AfD).[1] Also, radicalization on social media has been observed [17].

As a reaction to the massive fake news spread, most social networks have enabled methods to take care of the potential harms by curating, deleting and censoring content [18]. Thus, independent platforms now take the role of an information gatekeeper [19]. In 2018, the European Commission has appointed a "High Level Group on fake news and online disinformation". In October 2017, a German law came to force called *Netzwerkdurchsetzungsgesetz* (NetzDG; Network Enforcement Act). It attempts to fight fake news and hate speech by forcing platforms to quickly delete illegal contents but has been widely criticized for threatening freedom of speech.

2.2 Related Work

Over the last years, research on fake news has been emerging. Work focused on how fake news is received by individuals and potential factors determining whether fake news is considered true. Light has been shed on a cognitive heuristic called *Confirmation Bias* [20], describing the selective choice of (news) messages on the basis of matching ideologies. Marwick and Lewis [21] showed in several studies that individuals tend to forget the original source of information (*Sleeper Effect*). Subsequently, having forgotten about the untrustworthy source, the false information might be taken as true when fitting one's ideology [22]. Polage [23] described the *Truth Effect*, showing that content which is repeated often and by multiple sources tends to be taken for real at some point, independent from initial considerations.

A study by Del Vicario et al. [24] determined that (fake) news is spread especially within homogeneous groups of users with similar ideologies. In a Twitter analysis, Starbird [25] examined alternative narratives which she describes as a special kind of fake news, neither fitting one side of the conventional left-vs. right-wing cleavage but motivated by anti-globalism (e.g., criticism on the role of the US in world politics).

To investigate perception and handling of fake news by individuals, Barthel et al. [26] carried out a representative survey of 1,002 U.S. adults. According to the findings, about one third (32%) of U.S. adults "often see political news stories online that are made up" and 64% say fake news "cause[s] a great deal of confusion about the basic facts" [26]. Additionally, there are 23% who stated that they have shared a made-up story with 14% knowing it was fake. Furthermore, the participants were asked to state their opinion on responsibilities to prevent the spread of fake news. While 45% said government, politicians and elected officials have a great deal of responsibility, 43% held the public responsible, 42% social networking sites and search engines. The authors detected age as a determining factor: "Americans aged 50 and older are more likely to place a great deal of responsibility on the government" [26].

Allcott and Gentzkow [4] collected fake news headlines circulating during the U.S. election campaign. They asked 1,208 survey participants whether they had seen those headlines and whether they had initially accepted them as true. They found that 15% of

[1] https://correctiv.org/echtjetzt/artikel/2017/09/25/wahlcheck17-zieht-bilanz-den-fake-news-keine-chance/

the participants stated they had seen fake news headlines while 8% also believed them [4]. According to the authors, Democrats, heavy media consumers, people with higher education, and of higher age tend to “have more accurate beliefs about news” [4].

Sängerlaub [3] conducted a representative survey of 1,037 participants shortly after the German parliamentary elections in 2017. His results show that 61% of the participants had the impression that there were many cases of fake news during the election campaign. Interestingly, those who voted for the right-wing, populist party AfD agreed more to this statement than others (71%). Regarding fake news dealing with migration, right-wing supporters were often found to believe them. Besides partisanship, other factors like gender, age, income, and education had less influence on the results [3]. Lewandowsky et al. [27], [28] follow the assumption of ideology being a decisive factor, focusing on correlations between perceptions of fake news and worldviews. Reuter et al. found that 73% of the German population perceive false rumors in social media as a barrier for its use in emergencies [29].

Regarding the question of how to counteract fake news, Lewandowsky et al. [27] focused on building new narratives based on their examination of psychological dynamics with respect to the interaction with (fake) news. Other scholars introduced tools useful for detection and the individuals’ ability to assess news as untrustworthy [30–32]. A study by Neudert [33] on the issue of computational propaganda in Germany revealed a rise in junk news generated by social bots in the context of “the skeptical political zeitgeist” [33] in Europe. The authors further suggest junk news often to be disseminated by right-wing populist actors.

2.3 Research Gap

We aim for a deeper understanding of the highest populated and economically important European country with respect to fake news. While the country’s discourse is marked significantly less by fake news than debates around US elections are [34], it has, like many other societies, experienced a polarization of political attitudes in recent times, introducing the issue and providing fertile ground for the creation of fake news [35]. While a few studies have investigated fake news in Germany concerning specific election events [3], [34], there are no studies that analyzed the perception of fake news in Germany independently of specific events. With respect to possibilities of counteraction, we offer a first glance at a population’s general opinion on this issue. Reviewing related work, we propose to encounter the lack of research about citizens’ perceptions of countermeasures. These findings may contribute to the development of practical policies, as the success of ICT governance strongly depends on the willingness of non-state actors [36]. This may prove to be relevant as German decision-makers have been active in this field, introducing the Network Enforcement Act. We suggest our analysis to be helpful as survey answers are analyzed by both demographic and ideological characteristics, valuably pointing out different degrees of sensitivity in perception. Our findings concerning Germany’s population may not only be useful with respect to this single case but contribute to an understanding of fake news interaction in other European countries with similar media-cultural conditions and political polarization [34]. Instead of following US case studies and their binary ideological categorization of left- (Democratic) or right-wing (Republican) partisanship, we analyze the survey results according to demographic factors considered relevant by the surrounding academic discourse and test for correlations corresponding to ideological statements.

3 Approach

We intend to give an overview of the situation of fake news in Germany by finding answers to questions of attitude, interaction, and counteraction with respect to fake news. For these three topics, we also investigate differences between demographic subgroups and ask for correlations with more general attitudes regarding perceived threats to peace. To find answers to these questions, we conducted a representative survey of adults in Germany and analyzed the results. Using data of a representative survey, we compile a descriptive quantification about the belief of Germans.

3.1 Survey Design

We conducted a representative online survey (N=1023) of the adult German population in July 2017, using the ISO-certified panel provider GapFish (Berlin). They guarantee panel quality, data quality, and security, as well as survey quality through various (segmentation) measurements for each survey within their panel of 180,000 active participants. Our overall survey included 30 questions and also covered other topics, such as [37], [38]. In this work, we investigated three survey questions targeting the subject of fake news and selected on the basis of previous work. Each question consisted of multiple items which were answered on a three- or five-step Likert scale (see Appendix). The first question aimed at participants' attitudes towards fake news and their ramifications. The ten items comprised statements about fake news, which participants had to agree or disagree with on a five-step Likert scale. For example, the statements were about fake news posing a threat, or influencing public actors and population. In the second question, we asked whether and how participants had contact with fake news. On seven items, they had to state, among others, if they have perceived, liked, commented on, shared or created fake news on a 3-step Likert scale. Additionally, we asked for their opinion regarding the handling, i.e., counteracting of fake news. Participants were told to assess six suggested approaches to counteract fake news responding on a 5-step Likert scale. We selected several additional question items which we expected to hold interesting relations to fake news, including demographic positions and ideological statements.

3.2 Characteristics of Survey Participants

The conducted survey is representative of the German population stratified according to gender, age from 18 to 64 years, and federal state. This means that on these variables, the collected data shows no significant differences from the target population. Furthermore, we ensured a wide spread of the survey sample in terms of education and income which could not be tested for representativity due to the lack of appropriate national statistics. The χ^2-test on gender reveals no significant differences to the German population between 18 and 64 years (χ^2 (df= 1) = 0.031, p= .860). The collected sample consists of 50.4% male and 49.6% female participants (Germany: 50.7% male and 49.3% female) and, therefore, is representative. The participants' age was gathered in groups of 18 to 24 years, 25 to 34 years, 35 to 44 years, and 45 to 54 years. All group sizes correspond to the target population (χ^2 (df= 4) = 1.928, p= .749). We precluded one participant from analysis who stated to be 65 years or older, thus analyzing answers of 1,023 participants. Also, all sixteen German federal states are represented proportionately (χ^2 (df= 15) = 3.832, p= .998). Survey participants were paid a small allowance, recruited online, and able to take part once via the research agency's platform.

Most of all participants stated using a smartphone on a daily (49%) or hourly (43%) basis. Facebook is used less frequently than smartphones. A large majority (67%) does not use Twitter at all. Instagram is less popular than Facebook but more popular than Twitter, with 24% using it daily or hourly. All items concerning smartphone and social media use are highly influenced by the age of the participants (e.g., χ^2 (df= 4) = 195.2, p= .0000 for smartphone use), with younger respondents being more active.

3.3 Analysis

We examined the obtained survey data using R for data preparation and statistical methods, and Tableau for visual analytics and creation of data figures. For each question, we calculated the basic frequencies and created Diverging Stacked Bar Charts. Significant differences between participants of different demographic groups were determined using χ^2-tests of independence. Further, we investigated relationships to selected items of the extended question catalog. We calculated the Kendall rank correlation coefficient (Kendall's τ) for each pair of items, forming a correlation matrix and tested the correlations for non-zero coefficients.

4 Empirical Results

4.1 Attitude towards Fake News

The responses show that a large majority of participants concedes the risks of fake news (see Fig. 1). More than 80% agree that fake news poses a threat and can manipulate the population's opinion. But almost the same number of participants believes that public players like politicians can be manipulated as well. 78% see fake news harming democracy. In contrast, there are fewer (but still a significant amount of) people who see a threat in social bots and state censorship. Furthermore, there is a large number of participants who are indecisive about social bots. About 24% believe that fake news is annoying but does not pose a threat. There is no clear tendency, and there are widespread answers to the question of fake news being a pretext to fight system-critical actors. Most of the participants (82%) see the platform operators responsible for preventing fake news. 70% believe it is the state's task.

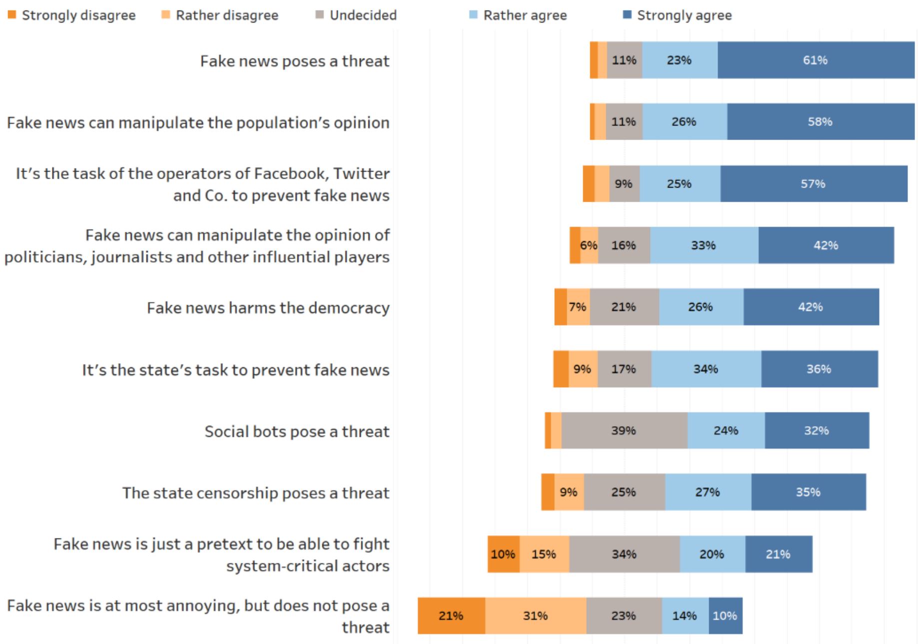


Fig. 1. Responses to the ten items of the first question ("Please indicate to what extent you agree to the following statements on fake news"). The responses are visualized in a *Diverging Stacked Bar Chart.* The level of agreement on an item is indicated by the horizontal shift.

Demographic Aspects. There are significant differences between female and male participants on several items: While about half of all women (47%) are undecided about a threat of social bots and 50% agree (or strongly agree) on it, male participants are more decisive (χ^2 (df= 4) = 34.168***[2], τ = -0.100***). 37% of male participants strongly agree (and 24% agree), and only 31% are undecided. There are also more male participants who disagree with it (7% male and 3% female). Regarding the threat of state censorship, men agree with this statement more strongly than women (41% vs. 29%, χ^2 (df= 4) = 25.394***, τ = -0.071*). Female participants clearly agree, that it is the state's task to prevent fake news (74%, disagree 9%, χ^2 (df= 4) = 21.029***, τ = 0.103***). Men are less consensual (65%) and more hostile (18%) to this statement. Most female participants are undecided (40%) whether fake news is just a pretext to fight system-critical actors (χ^2 (df= 4) = 24.366***, τ = 0.031). Male participants are less undecided (29%) and more disagreeing (30% vs. female 20%). However, there are fewer female participants who strongly agree (18% vs. male 23%) and more who just agree than men (22% vs. 18%). Furthermore, there are significant differences between different age groups on this item (χ^2 (df= 16) = 33.63**, τ = 0.065**). Older participants are less undecided than younger ones (e.g., 30% of 55 to 64 years vs. 40% of 25 to 34 years). The group of 55 to 64 years is also more disagreeing on this item than others (34% vs.

[2] (p<.001)***;(p<.01)**; (p<.05)*

15% of 18 to 24 years). In addition, the youngest group (18 to 24 years) is more agreeing than others (51% vs. e.g., 37% of 25 to 34 years).

The educational level has an important impact on many questions. For example, on the item of social bots (χ^2 (df= 20) = 65.714***, τ = 0.095***), participants with a university degree or university of applied sciences degree are mainly strongly agreeing to a threat of social bots (43%). However, participants with a high school diploma or secondary school degree are mainly undecided (38%, 47%). The item of social bots is also influenced by income (χ^2 (df= 12) = 23.864*, τ = 0.062*) which slightly correlates with educational level: Kendall's τ = .0748**. The majority of those with an income of more than €3,500 strongly agree that social bots pose a threat. Other income groups are mainly undecided (38% to 41%).

Correlations with Additional Survey Items. The correlation matrix generally revealed a high number of significant relationships while we consider only the most significant and compelling links to other survey items. We found that the question about the assessment of social media advice had significant non-zero coefficients especially for the item of correcting wrong information ("When you find or share wrong information, correct it."). This item correlates to multiple items about fake news: For example, the response on this item has influence on responses towards fake news posing a threat for democracy (τ = .21***, z = 7.87). There also is a quite strong connection to stating that fake news manipulates the population's opinion (τ = .31***, z = 11.10). A smaller (but still significant) link can be found regarding fake news manipulating the opinion of politicians, journalists and other influential players (τ = .23***, z = 8.65). The item stating that "all information controlled by the state should be made publicly available, even if posing a risk to public safety" shows a correlation with the attitude towards fake news being at most annoying and not posing a threat (τ = .21***, z = 8.45). Similarly, they rather disagree with the item stating that fake news poses a threat (τ = −.13***, z = −4.93) and they more often think that fake news is just a pretext to fight system-critical actors (τ = .18***, z = 7.20). We could also find a negative link to the item of fake news manipulating the population's opinion (τ = −.15***, z = −5.58) and another negative coherence regarding the responsibility of social media operators to prevent fake news (τ = −.11***, z = −4.10). On the items about what poses a threat to the peace in Germany, the analysis revealed some significant correlations. Seeing other states as a threat to peace is positively correlated to assessing fake news posing a threat (τ = .12***, z = 4.65) and harming democracy (τ = .10***, z = 3.84). We found a similar connection to the threat of social bots (τ = .10***, z = 3.75). Interestingly, a threat of social bots correlates with seeing a threat in nationalism (τ = .19***, z = 7.29). The items on the threat of nationalism and capitalism also show a significantly positive coefficient on fake news harming democracy (τ = .21, ***, z = 7.82 and τ = .10***, z = 3.82). There also might be an opposing link between the item of multicultural coexistence posing a threat to peace and the item regarding a threat of fake news (τ = −.08**, z = −3.06). Furthermore, the analysis revealed correlations with the statement that fake news is just a pretext to be able to fight system-critical actors: There are significantly positive coefficients for the item on a threat of globalization (τ = .21***, z = 8.14) and other states (τ = .19***, z = 7.50).

4.2 Interaction with Fake News

While almost half of all participants (48%) stated they had perceived fake news, the majority says they had not interacted with fake news in any of the suggested ways (see figure 2). In general, the number of participants who affirm interaction with fake news is between 23% ("deleted or reported fake news") and 2% ("created fake news"). The responses of *only* those participants who have perceived fake news are slightly higher: There are 38% saying they have deleted or reported fake news and 23% who affirmed commenting and disliking fake news, respectively. About 7% state they had shared fake news and 4% say they had created fake news themselves. However, due to the social desirability bias, these numbers are probably not accurate and might be higher. Since fake news has a negative social standing, not everyone would respond honestly and admit e.g., sharing or creating fake news.

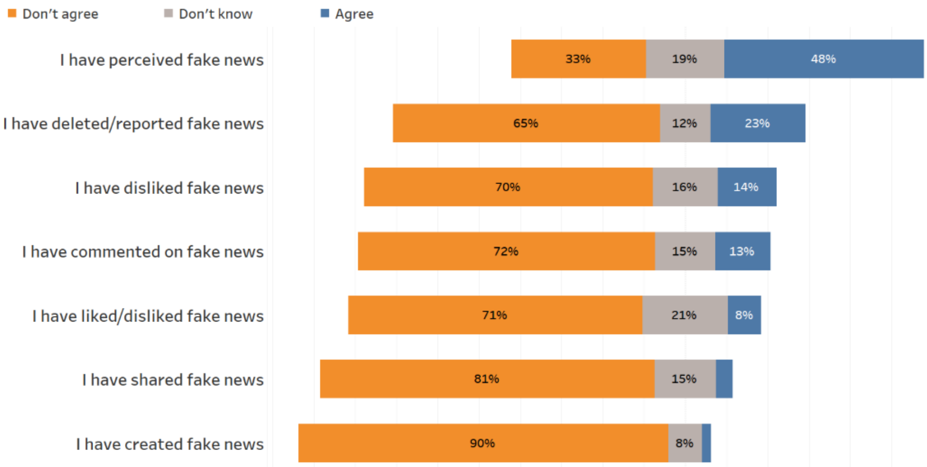


Fig. 2. Responses to the seven items of the second question ("How did you have contact with fake news on social networks?").

Demographic Aspects. Regarding gender, there are no significant differences in interaction with fake news. However, age has an important impact on most questions. For example, 63% of the participants aged 18 to 24 and 59% aged 25 to 34 say they have perceived fake news while only 33% of participants aged 55 to 64 agree (χ^2 (df= 8) = 52.356***, τ = 0.140***). Additionally, younger participants are more likely to have liked, disliked (τ = 0.021) or commented (τ = 0.044) on fake news. Furthermore, younger participants stated significantly more often than older participants to have deleted or reported fake news (46% aged 18 to 24 vs. 11% aged 55 to 64, χ^2 (df= 8) = 78.881*** τ = 0.149***). The educational level corresponds significantly with the perception of fake news (χ^2 (df= 8) = 37.012***, τ = 0.141***, ignoring the minority without graduation) but not with the interaction with it. 67% of participants with a university degree state to have perceived fake news while only 41% of participants with a secondary degree agree. A significant impact of income on the interaction with or perception of fake news could not be shown.

Correlations with Additional Survey Items. Again, we found several links to additional survey items using rank-based correlation. Unsurprisingly, items on social media usage show a correlation with the interaction with fake news. For example, posting messages on social media correlates with commenting on (τ = .22***, z = 7.88), sharing (τ = .12***, z = 4.29) and reporting (τ = .20***, z = 7.30) fake news. Thus, a higher activity in social networks often comes along with more interaction with fake news. A possible (but weak) indicator for the item on fake news creation is the belief that globalization is a threat to peace (τ = .08**, z = 2.92). Responses to the item on public safety having priority over information access show a minor correlation to the item about perceiving fake news (τ = −.06*, z = −2.19). In contrast, the item on unrestricted information access shows a significantly positive coefficient to the item on fake news creation (τ = .13***, z = 4.57).

4.3 Counteracting Fake News

Most participants agree with all suggested ways to deal with fake news (see: figure 3). For all items, the numbers of participants who agree vary between 80% ("quick reactions of the authorities") and 72% ("establish state IT centers of defense"). The amount of neutral responses ranges from 14% to 21%, while very few participants do not agree with the suggested approaches (between 3% and 7%). Compared to the other items, the idea of a state IT center managing fake news has the lowest compliance. It has the lowest number of supporting answers and the highest numbers of neutral and disagreeing answers. However, the reported deviations are altogether relatively small. For gender, age, educational level, and income, no significant differences could be found in the answering patterns.

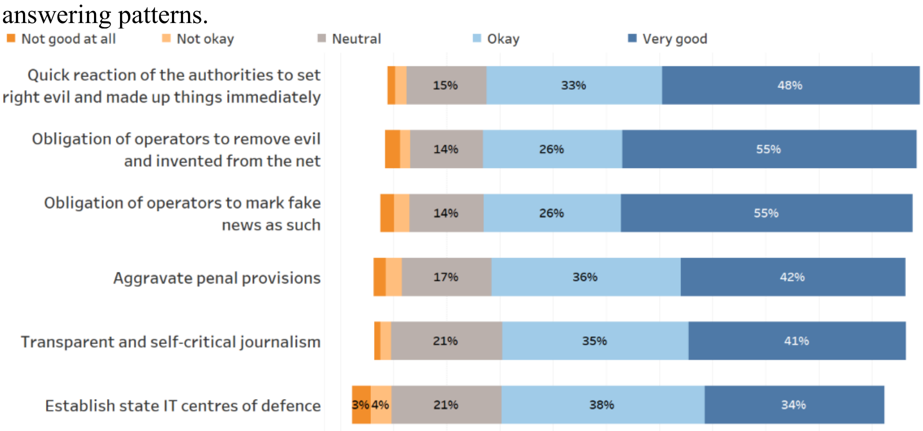


Fig. 3. Responses to the six items of the third question: ("How do you rate the following suggestions for dealing with fake news?"), visualized in a Diverging Stacked Bar Chart.

Correlations with Additional Survey Items. The correlation revealed significant connections to the items on an obligation of operators to remove fake news and to mark fake news as such. The social media advice on correcting false information has positive coefficients towards obligation to remove fake news (τ = .36***, z = 13.11) and to mark fake news ($\tau = .34^{***}$, $z = 12.38$). The same links can be found for the advice to verify information before publishing (obligation to delete fake news: $\tau = .41^{***}$, $z = 14.65$ and obligation to mark fake news: $\tau = .39^{***}$, $z = 13.89$).

5 Discussion and Conclusion

5.1 Scientific Contribution & Policy Implications

In sum, the analysis has shown that a large majority of participants concedes the risks of fake news. More than 80% of the participants agree that fake news poses a threat, and 78% see fake news as harming democracy. Even though almost half of all participants (48%) stated that they had perceived fake news, the majority claims not to have interacted with them in any of the suggested ways. Only 2% admitted having created fake news. Moreover, most participants agreed with all suggested ways to deal with or counteract fake news.

We analyzed survey results with respect to common demographic factors. While studies did not make any differences between male and female respondents, we point out to respective diverging attitudes [39], with women stressing the state's responsibility to deal with fake news and men increasingly pointing out to the danger of state censorship. Our results propose future policy-making to be sensitive towards this issue by ensuring to enhance a more differentiated view considering *gender*. Yet, one may put this into perspective, as our findings do not suggest a correlation between gender and any item regarding interaction or counteraction with respect to fake news. The same applies for *income* as a potential factor. Further, we can assume only the highest income group, slightly correlating with education, to be critical of social bots. Thus, to allow for overall informed opinions and decisions regarding fake news, it may be useful to raise awareness to this phenomenon. In contrast, *age* and *education* influence not only attitudes regarding fake news but also people's considerations of interaction with them. Our findings support the view of younger and relatively educated people being more informed and sensitive about fake news due to their relatively frequent and differentiated internet usage, contradicting other studies [4]. Thus, educational approaches may specifically target older generations and focus on easy and appealing access to necessary information [40]. Yet, it should be kept in mind that across correlations between demographic factors and survey items, we could mostly make out only small effects.

We analyzed respondents' answers according to ideological motives, testing for correlations with statements reflecting perceived threats. Respondents who believed globalization to pose a threat to peace in Germany also shared the attitude of perceiving the phenomenon of fake news used as a pretext to fight system-critical actors and showed a weak link to the action of creating fake news. Following Starbird's assumption of online anti-globalist narratives being often (re-)produced by nationalist populist groups [25] and the findings of Allcott et al. [4] pointing out to right-wing or republican ideology as a motivation for creating fake news, we suggest both the creation and playing down of fake news to potentially symbolize the respective populist political view. Similarly, respondents who possibly downplayed fake news to be no threat and just an excuse to fight against system-critical actors perceived multicultural coexistence to be threatening. This fits with Sängerlaub's result on AfD-related appreciation of fake news [3]. Future regulative implementations might, therefore, reflect awareness of diverging attitudes across political camps. Our findings also indicate German respondents to not associate fake news with mainstream media as apparently as in the US, diverging from study results focused on the immediate aftermath of elections [3]. Respondents who perceived nationalism as well as capitalism to pose a threat also found fake news to harm democracy while the first group also perceived social bots to be problematic, suggesting relatively left-wing or liberal respondents to be more critical of fake news. Assuming a perceived dualism of freedom and security, respondents who prioritized unrestricted access to information over public safety shared the attitude of fake news not posing a threat and legitimizing actions against system-critical actors as well as showed to be responsible for the creation of fake news. Participants who prioritized public safety demanded control with respect to fake news by social media operators. Policies may consider this regarding burden-sharing. Again, effects between ideological standpoints and survey items were not great, yet, stronger relationships compared to demographic factors and items were proposed.

5.2 Limitations & Future Work

This work has limitations: (1) Our results were acquired using a survey. This method of data collection implies the risk of participants being more technophile, not responding honestly, being influenced by social desirability biases, which is very likely on such a controversial topic. To find more reliable numbers, other techniques than a survey need to be utilized. (2) Furthermore, the study relies on self-reported behavior. It is likely that people interact with fake news without knowing it. (3) Also, it must be mentioned that the concept of fake news – as a result of the term's ambiguous nature and changing meaning in the current political and societal discourse – might have been interpreted differently by each survey participants. (4) Also, the correlations to other survey items are all lower than $\tau = .5$ and thus, cannot be considered strong. However, although the given correlations are weak, they are statistically significant and reveal existing tendencies. (5) Our panel excluded teenagers younger than 18 years; a group which might have yielded important results regarding the perception of fake news. Future research may test for correlations between the three respective topics to gain deeper insight into causal relationships between general perception, perceived interaction and opinion on counteraction while targeting various groups defined by demographic and ideological factors. Worldview-related variables may contribute to a more accurate understanding of variations in perceptions regarding fake news and of a diversity of fake news conceptions.

Acknowledgements: This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP, the BMBF within KontiKat (13N14351) and the German Research Foundation (DFG) within the CRC 1119 CROSSING. We like to thank Sabrina Neuhäusel for her corrections of the paper.

References

1. Reuter, C.: Information Technology for Peace and Security - IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace. Springer Vieweg, Wiesbaden, Germany (2019)
2. Reuter, C.: Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement. Springer Vieweg, Wiesbaden (2018). http://www.springer.com/de/book/9783658195229
3. Sängerlaub, A.: Verzerrte Realitäten: „Fake News“ im Schatten der USA und der Bundestagswahl. Stiftung Neue Verantwortung, Berlin, Germany (2017). https://www.stiftung-nv.de/de/publikation/verzerrte-realitaeten-fake-news-im-schatten-der-usa-und-der-bundestagswahl
4. Allcott, H., Gentzkow, M.: Social Media and Fake News in the 2016 Election. J. Econ. Perspect. 31, 211–236 (2017)
5. Kaufhold, M.-A., Reuter, C.: Konzept und Evaluation einer Facebook-Applikation zur crossmedialen Selbstorganisation freiwilliger Helfer. In: Thomas, O. and Teuteberg, F. (eds.) Proc. WI (WI). pp. 1844–1858. University of Osnabrück, Osnabrück, Germany (2015)
6. Reuter, C., Pätsch, K., Runft, E.: Terrorbekämpfung mithilfe sozialer Medien – ein explorativer Einblick am Beispiel von Twitter. In: Leimeister, J.M. and Brenner, W. (eds.) Proc. WI (WI). pp. 649–663. University of St. Gallen, St. Gallen, Switzerland (2017)
7. Reuter, C., Kaufhold, M.-A.: Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics. J. Contingencies Cris. Manag. 26, 41–57 (2018)
8. European Comission: A Multi-Dimensional Approach to Disinformation. (2018)
9. Becker, B.W.: The Librarian’s Information War. Behav. Soc. Sci. Librar. 35, 188–191 (2016)
10. 10. Cooke, N.A.: Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age. Libr. Q. 87, 211–221 (2017)
11. Wardle, C., Derakhshan, H.: Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe, Strasbourg, France (2017). https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c
12. Ziegele, M., Breiner, T., Quiring, O.: What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. J. Commun. 64, 1111–1138 (2014)
13. Stieglitz, S., Dang-Xuan, L.: Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. J. Manag. Inf. Syst. 29, 217–248 (2013)
14. Perrin, A.: Social Media Usage: 2005 - 2015, http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/
15. Sängerlaub, A.: Deutschland vor der Bundestagswahl: berall Fake News?! Stiftung Neue Verantwortung, Berlin (2017). https://www.stiftung-nv.de/sites/default/files/fakenews.pdf
16. Flanagin, A.J., Metzger, M.J.: Digital Media and Youth: Unparalleled Opportunity and Unprecedented Responsibility. In: Metzger, M.J. and Flanagin, A.J. (eds.) Digital media, youth, and credibility. pp. 5–28. MIT Press, Cambridge, MA (2008)
17. Riebe, T., Pätsch, K., Kaufhold, M.-A., Reuter, C.: From Conspiracies to Insults: A Case Study of Radicalisation in Social Media Discourse. In: Mensch und Computer 2018: Workshopband. pp. 595–603. Dresden, Germany, Germany (2018)
18. Tene, Omer; Polonetsky, Jules; Sadeghi, A.-R.: Five Freedoms for the Homo Deus. IEEE Secur. Priv. 16, 15–17 (2018)

19. Wohn, D.Y., Fiesler, C., Hemphill, L., De Choudhury, M., Matias, J.N.: How to Handle Online Risks? In: Proc. CHI EA. pp. 1271–1276. ACM Press, New York, (2017)
20. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2, 175–220 (1998)
21. Marwick, A., Lewis, R.: Media manipulation and disinformation online, https://datasociety.net/output/media-manipulation-and-disinfo-online/
22. Kumkale, G.T., Albarracín, D.: The sleeper effect in persuasion: a meta-analytic review. Psychol. Bull. 130, 143–72 (2004)
23. 23. Polage, D.C.: Making up history: false memories of fake news stories. Eur. J. Psychol. 8, 245 (2012)
24. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., Quattrociocchi, W.: The spreading of misinformation online. Proc. Natl. Acad. Sci. U. S. A. 113, 554–559 (2016)
25. Starbird, K.: Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In: Proc. Web and Social Media. pp. 230–239. AAAI Press, Palo Alto, CA (2017)
26. Barthel, M., Mitchell, A., Holcomb, J.: Many Americans Believe Fake News Is Sowing Confusion, http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/
27. Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N., Cook, J.: Misinformation and Its Correction: Continued Influence and Successful Debiasing. Psychol. Sci. Public Interes. 13, 106–131 (2012)
28. Lewandowsky, S., Stritzke, W.G.K., Freund, A.M., Oberauer, K., Krueger, J.I.: Misinformation, disinformation, and violent conflict: From Iraq and the “War on Terror” to future threats to peace. Am. Psychol. 68, 487–501 (2013)
29. Reuter, C., Kaufhold, M.-A., Spielhofer, T., Hahne, A.S.: Social Media in Emergencies: A Representative Study on Citizens’ Perception in Germany. Proc. ACM Hum. Comput. Interact. Comput. Coop. Work Soc. Comput. 1, 1–19 (2017)
30. Saez-Trumper, D.: Fake Tweet Buster: A Webtool to identify Users Promoting Fake News Ontwitter. In: Proc. Hypertext and Social media. pp. 316–317. ACM Press, NY (2014)
31. Narwal, V., Salih, M.H., Lopez, J.A., Ortega, A., O’Donovan, J., Höllerer, T., Savage, S.: Automated Assistants to Identify and Prompt Action on Visual News Bias. In: Proc. CHI Extended Abstracts. pp. 2796–2801. ACM Press, New York, NY (2017)
32. Jaho, E., Tzoannos, E., Papadopoulos, A., Sarris, N.: Alethiometer: A Framework for Assessing Trustworthiness and Content Validity in Social Media. In: Proc. WWW ’14 Companion. pp. 749–752. ACM Press, New York, NY (2014)
33. Neudert, L.-M.: Computational Propaganda in Germany: A Cautionary Tale. Project on Computational Propaganda, Oxford, UK (2017). http://blogs.oii.ox.ac.uk/politicalbots/wp-content/uploads/sites/89/2017/06/Comprop-Germany.pdf
34. Neudert, L.-M., Kollanyi, B., Howard, P.N.: Junk News and Bots during the German Parliamentary Election: What are German Voters Sharing over Twitter? Data Memo 7. Project on Computational Propaganda, Oxford, UK (2017). https://comprop.oii.ox.ac.uk/research/junk-news-and-bots-during-the-german-parliamentary-election-what-are-german-voters-sharing-over-twitter/
35. Caiani, M., Kröll, P.: Nationalism and Populism in Radical Right Discourses in Italy and Germany. Javn. - Public. 24, 336–354 (2017)
36. Löblich, M., Wendelin, M.: ICT policy activism on a national level: Ideas, resources and strategies of German civil society in governance processes. New Media Soc. 14, 899–915 (2012)

37. Kaufhold, M.-A., Gizikis, A., Reuter, C., Habdank, M., Grinko, M.: Avoiding Chaotic Use of Social Media during Emergencies: Evaluation of Citizens' Guidelines. J. Contingencies Cris. Manag. 1–16 (2019)
38. Kaufhold, M.-A., Grinko, M., Reuter, C., Schorch, M., Langer, A., Hollick, M., Skudelny, S., Hollick, M.: Potentiale von IKT beim Ausfall kritischer Infrastrukturen: Erwartungen, Informationsgewinnung und Mediennutzung der Zivilbevölkerung in Deutschland. In: Proc. WI. pp. 1–15. AIS, Siegen, Germany, Germany (2019)
39. Baumann, A., Krasnova, H., Veltri, N.F., Yunsi, Y.: Men, Women, Microblogging: Where Do We Stand? In: Proc. WI. pp. 857–871. University of Osnabrück, Osnabrück (2015)
Mirsch, T., Lehrer, C., Jung, R.: Digital Nudging: Altering User Behavior in Digital Environments. In: Proc. WI. pp. 634–648. St. Gallen, Switzerland (2017)

Appendix: Survey Questions

#15) Attitude towards Fake News. Please indicate to what extent you agree to the following statements on fake news. (strongly agree, rather agree, undecided, rather disagree, strongly disagree): *1. Fake news poses a threat. | 2. Social bots pose a threat. | 3. It's the state's task to prevent fake news. | 4. It's the task of the operators of Facebook, Twitter and Co. to prevent fake news. | 5 Fake news harms the democracy. | 6. Fake news can manipulate the opinion of politicians, journalists and other influential players. | 7. Fake news can manipulate the population's opinion. | 8. Fake news is just a pretext to be able to fight system-critical actors. | 9. The state censorship poses a threat. | 10. Fake news is at most annoying but does not pose a threat.*
#16) Interaction with Fake News. How did you have contact with fake news on social networks? (agree, don't know, don't agree): *1. I have perceived fake news | 2. I have liked/disliked fake news | 3. I have commented on fake news | 4. I have shared fake news | 5. I have deleted/reported fake news | 6. I have disliked fake news*
#17) Dealing with Fake News. How do you rate the following suggestions for dealing with fake news? (very good, okay, neutral, not okay, not good at all): *1. Establish state IT centers of defense | 2. Aggravate penal provisions | 3. Transparent and self-critical journalism | 4. Quick reaction of the authorities to set right evil and made up things immediately | 5. Obligation of operators to remove evil and invented from the net | 6. Obligation of operators to mark fake news as such*
Further questions:
#1) Please state how often you perform the following activities. (hourly, daily, more than once a week but less than daily, less than once a week, never): *1. Using a smartphone | 2. Using Facebook | 3. Using Twitter | 4. Using YouTube | 5. Using other forms of social media | 6. Posting messages to social media*
#13) In your opinion, should federal authorities in Germany have the right or not to guarantee public safety? (definitely yes, rather yes, neutral, rather no, definitely no): *1. All information controlled by the state should be made publicly accessible, even if posing a risk to public safety. | 2. The public safety should have priority, even if this restricts the access to information controlled by the state.*
#18) In your opinion, what poses the greatest threat to peace in Germany? (very great, great, neutral, little, very little): *1. Other States | 2. Nationalism | 3. Multicultural coexistence | 4. Climate change | 5. Social injustice | 6. German politics | 7. Left-wing extremism | 8. Religious fanaticism | 9. Capitalism | 10. Globalization*
#23) How do you assess the following advice for using social media during crisis situations? (very important, rather important, neutral, rather unimportant, very unimportant): *1. When you find or share wrong information, correct them.*
#25) How do you assess the following aspects of social media usage during crisis situations? (strongly agree, agree, neutral, disagree, strongly disagree): *1. You are responsible for your postings, please consider possible consequences. | 2. Verify your information before publishing.*

Analyzing the Potential of Graphical Building Information for Fire Emergency Responses: Findings from a Controlled Experiment

Julian Weidinger[1], Sebastian Schlauderer[1], and Sven Overhage[1]

[1] University of Bamberg, Chair of Industrial Information Systems, Bamberg, Germany
{julian.weidinger,sebastian.schlauderer,sven.overhage}
@uni-bamberg.de

Abstract. To better support firefighters during emergency response processes, novel information technologies are frequently being presented in research and practice. While such approaches are often technology-driven in nature, we present a task-centered approach to identify the actual information demand during emergency response scenarios. As an important example, we examine the search and rescue task. Combining the theory of situation awareness with findings from cognitive science, we hypothesize that providing graphical information about the building and the location of victims increases firefighters' task performance in comparison to a verbal briefing. Findings from a controlled experiment that was developed in cooperation with a state firefighting academy show that such information might indeed facilitate the task performance. A continuous access to such information during the entire mission was found to be less effective, though. Our findings have implications for the development of novel information technologies and call for an adaption of current working routines.

Keywords: Firefighter information technologies, Situation awareness, Cognitive Science, Laboratory experiment.

1 Introduction

Newly emerging firefighter information technologies (FITs) such as digital plans [1, 2], unmanned aerial vehicles [3, 4], on-site emergency response systems [5, 6], augmented reality devices [7, 8], or intelligent protective clothing [9, 10] open up novel opportunities to gather, process, and present real-time information about the site of a fire emergency. They are hence supposed to bear a significant potential to facilitate the making of context-dependent decisions and, accordingly, to support the emergency response process. Currently, however, novel FITs are developed and proposed mainly based on their innovative capabilities and the presumably resulting potential to support the emergency response process. Recent studies indicate that such technology-driven approaches run a risk to neglect the specific nature of emergency response processes and might miss the actual demand for information [11].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

During emergency responses, firefighters make time-critical decisions in a dynamic environment. By improving the information base, new and emerging information technologies might facilitate the decision-making process. However, there also exist tight constraints, for instance with respect to the time and the margin for error. The provided information hence needs to be succinct, easy to process and straightforwardly understandable without distraction. To make sure that firefighter information technologies indeed meet the demand that exists on site, it appears thus necessary to investigate, (i) which kind of information is effective in enhancing the performance of firefighters during their operations and (ii) how information technologies should be designed to deliver the information adequately. Yet, so far literature hardly discusses under which circumstances (and subject to which task-specific requirements) novel FITs are viewed as beneficial by prospective users.

With the work at hand, we intend to contribute to the closure of this literature gap. Narrowing the scope, we focus on examining, which kind of information is effective in supporting the search and rescue task. This task was identified as an appropriate study object during a roundtable discussion with instructors of a firefighting academy for several reasons: first, the task frequently occurs during daily emergency responses. Second, firefighters typically suffer from an insufficient information basis during this task. Third, emerging FITs like digital plans and on-site emergency response systems are proposed in literature to better support this task (among others). In our study, we examine the following research questions: "*Does an up-front presentation of a graphical building plan and the location of a victim during the mission briefing improve the search and rescue task compared to a verbal description? Does the continuous availability of a graphical plan during the mission further improve the task?*"

To achieve rigorous results, we conducted a controlled experiment, in which 69 firefighter squads were provided with information about the building and the victim location in different formats and with different forms of availability. The experiment was conducted in cooperation with a Bavarian firefighting academy, which made available a building that has been designed for fire emergency response training purposes and resembles a typical small apartment house in Germany. The design of the experiment is informed by findings from cognitive science on the processing of information formats and on cognitive overload. The experiment results indicate in how far graphical building information is effective in enhancing typical fire emergency response operations. Accordingly, they provide an empirically grounded basis for the design of firefighter information technologies such as digital plans, on-site emergency response systems, or augmented reality devices, which shall make available such information to firefighters during the emergency response. As domains with high stress levels, strict timing constraints, and low margins for error hardly have been in the focus of research so far, the results of our study furthermore complement existing findings from cognitive sciences.

In section 2, we discuss the background of our study. Furthermore, we develop the hypotheses and the research model underlying the design of our experiment. In section 3, we describe the experiment design. Section 4 presents the obtained results. We discuss the results and the implications for academia and practice in section 5. In section 6, we conclude by summarizing the findings and outlining future research directions.

2 Background, Hypotheses, and Research Model

During fire emergency response operations, firefighters make time-critical decisions in highly dynamic situations. Because of the existing time constraints, they often must decide on a basis of insufficient information. Improving the information basis and increasing the so-called situation awareness is hence supposed to be a critical success factor to achieve better and/or quicker decisions [12, 13]. The theory of situation awareness can be applied to various domains, in which actors make time-critical decisions in stressful situations. Besides emergency responders like firefighters, it has also been applied to pilots and soldiers. Generally, situation awareness is established in a three-level process of perception, comprehension, and projection [12]. In the perception phase, the actor must capture all relevant aspects of the situation. In a simplified example, a firefighter would have to realize that there is smoke coming from an open window on the second floor or that the house door is closed. In the comprehension phase, the actor must try to connect the different aspects to build a holistic understanding of the situation. The firefighter would have to understand that there is a fire, which could be reached by breaking the door or using a ladder to reach the window. Finally, based on perception and comprehension, the actor can make projections about the near future. The firefighter could predict that the fire might spread to other rooms and could be reached faster through the window. All three levels of situation awareness demand qualitatively and quantitatively sufficient information. In this context, information technologies such as the FITs mentioned in the previous section can be an effective means to deliver this information. Looking at the research conducted on situation awareness in emergency response management, however, there is a strong focus on large-scale phenomena like the handling of extraordinary disasters or the coordination of multiple agencies [14, 15]. To our best knowledge, the theory has not yet been applied to study daily routines of firefighters such as search and rescue of victims.

2.1 Development of Hypotheses

To identify how information needs to be presented in such a scenario to increase situation awareness, we consulted relevant literature of cognitive science. Various studies in this field indicate that especially graphical information can be used to effectively and understandably provide information to humans. The graphical format often proved to be superior to other forms of presentation like natural language or written words. Generally, humans seem to better recall pictures and other graphical information than the identical information coded in words [16-18]. As an explanation for this, Kaplan, Kaplan and Sampson [16] showed that the human brain double-codes pictures (i.e., both verbally and visually). This mental connection between graphics and words enables people to better recall the information. A special form of graphical information are diagrams. Due to the support of numerous perceptual interferences and the grouping of related information, they can further improve comprehensibility [19]. Mousavi, Low and Sweller [20] showed that another improvement can be achieved by connecting the graphical information with corresponding verbal explanations. According to Mayer and Sims [21], these effects specifically apply to inexperienced persons. Since firefighters typically are not acquainted with the concrete situation and environmental conditions when responding to an emergency (in our scenario the apartment they will be operating in), they can also be characterized as inexperienced. Deducing from the above findings of cognitive science, we assume that firefighters with access to graphical information about a building shall be able to better recall this information during a mission. Accordingly, they should also be able to perform search and rescue tasks better. With our study, we apply the general findings of cognitive science to the practical and very special domain of firefighters. Firefighters typically work under high emotional stress and time pressure in chaotic situations. Yet, according to instructors of the firefighting academy who supported this study, the use of graphical information during responses has by now hardly been identified as a means to overcome these problems in practice. By applying existing findings of cognitive science to this special and comparably unexplored domain, we hence test (and might further increase) their general validity.

The specific characteristics of the firefighter domain, however, also imply a risk that firefighters' cognitive abilities might not suffice to process all information they are receiving. This problem is referred to as cognitive overload [22]. In an emergency, responders are confronted with countless environmental and emotional impressions that need to be processed. Once reaching the point of cognitive overload, they may not be able to completely perceive and memorize information like the graphical building plan examined in our study. For these reasons, we assume that making graphical information available continuously during the mission further enhances the task performance as the information would not have to be memorized but could be looked up when needed. We hypothesize that such a higher degree (of availability) of graphical information will have a positive effect on its usability and, accordingly, the task performance.

2.2 Research Model

To capture the background of our study more precisely, we integrate our assumptions into a research model. We postulate that different degrees of graphical information provide differing support for firefighters during their tasks. Particularly, we expect the different degrees of information to have an impact on their usability for firefighters. Usability as such can be further operationalized into three aspects: the effectiveness (i.e., the accuracy with which a task is fulfilled), the efficiency (i.e., the effectiveness in relation to the effort needed to complete a task), and the satisfaction (i.e., the users' comfort while performing a task) [23, 24]. Following this definition, we concretize our three hypotheses as follows (the resulting research model is summarized in Figure 1):

- **H_1**: Firefighters provided with a higher degree of graphical building information will be able to perform the task more accurately.
- **H_2**: Firefighters provided with a higher degree of graphical building information will need less time to perform the task accurately.
- **H_3**: Firefighters provided with a higher degree of graphical building information will be more satisfied while performing the task.

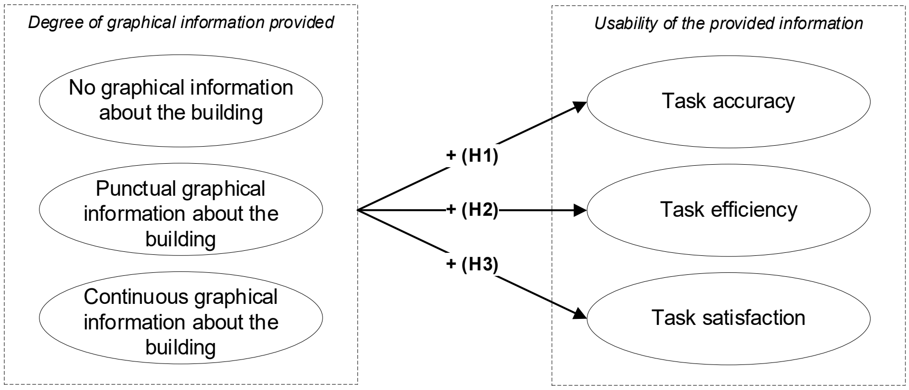


Figure 1. Research model

3 Research Procedure

To examine H_1-H_3, we conducted a controlled laboratory experiment. Next, we present the design of this experiment to evaluate H_1 and H_2. Thereafter, we describe the questionnaire that accompanied the experiment and was used to examine H_3.

3.1 Controlled Laboratory Experiment

Our experiment was developed and conducted in cooperation with the Bavarian state firefighting academy in Würzburg, Germany. In this academy, firefighters from all over Bavaria are trained in several tasks like salvage rescue, firefighting, and incident command. By means of the academy's specialized training facilities, we were able to simulate a realistic scenario under safe conditions. Despite providing a realistic scenario, the environment nevertheless allowed us to control surrounding factors in order to conduct a laboratory experiment following the guidelines given by Cox and Reid [25].

The experimental setting was developed in a collaboration with a group of instructors of the firefighting academy. Beginning with a roundtable session in late 2017, we identified various everyday tasks in which firefighters typically suffer from a lack of information. As one of the most critical of these scenarios, the instructors introduced the search and rescue of a victim during an apartment or house fire. Based on this scenario, we developed our concrete experiment in a second roundtable session.

In our experiment, a squad of firefighters responded to an apartment fire in the basement of a two-story building. The building was represented by the academy's fire training building (cf. Figure 2). It emulates a typical residential building and was specifically designed to simulate apartment fires by means of remotely controlled gas-fires and simulated smoke from fog machines. Inside the burning apartment, a life-size dummy depicted a missing, unconscious person. The squad responding to the apartment fire consisted of a squad leader and a team of two firefighters. The squad leader commanded the response by first investigating the scene and then briefing the team. In the experiment, the squad leader was represented by a member of our research team. In so doing, we could control the equality of the briefings. The team's performance, consequently, was not influenced by the abilities of its squad leader but only by the received information. The firefighter team got a briefing on the situation and was then commanded to enter the burning apartment to search and rescue the missing person. The positions of the team were taken by random firefighters wearing self-contained breathing apparatuses (SCBA). Other positions of a real-life squad like the operator of the fire engine and a second team of firefighters on standby have not been included in the experiment, because they would have had no connection to our experimental interventions.

The subjects who acted as team members were retrieved from firefighting courses, which took place from March to July 2018. Due to capacity restrictions, places in those courses are limited and distributed among all Bavarian fire departments. This resulted in a diverse, random sample of subjects. Each team consisted of two subjects and was randomly assigned to one of three equally sized groups [25]: G_0, G_1, and G_2. $\mathbf{G_0}$ was the control group. Teams of this group received a verbal briefing and command without any graphical resources, which represents the current state of the art for such rescue missions. The teams were briefed with the following standardized phrase: "There's an apartment fire in the basement, one person is missing but we know their presumable location. Heading down the stairs, there are four doors: one ahead, two on your left, and one at your left back. To reach the person, you must take the farer left door. Then cross the following corridor straight ahead. In the room after the corridor, you must go to the right around a room divider. There, you should find the person. Intervention team for search and rescue with the first hose line to the basement via the stairs, go!"

$\mathbf{G_1}$ was the first treatment group. As treatment, teams of this group received punctual access to a sketch-map of the apartment during the briefing. As shown in Figure 2, the map contained the apartment's layout including stairs, walls, doors and the victim's location. The squad leader briefed them with the identical phrase that was also used in the control group. This time, however, he visualized the verbal information in the map, for example by pointing toward the doors the team was supposed to go through. After the briefing, the team entered the building without any further access to the sketch-map.

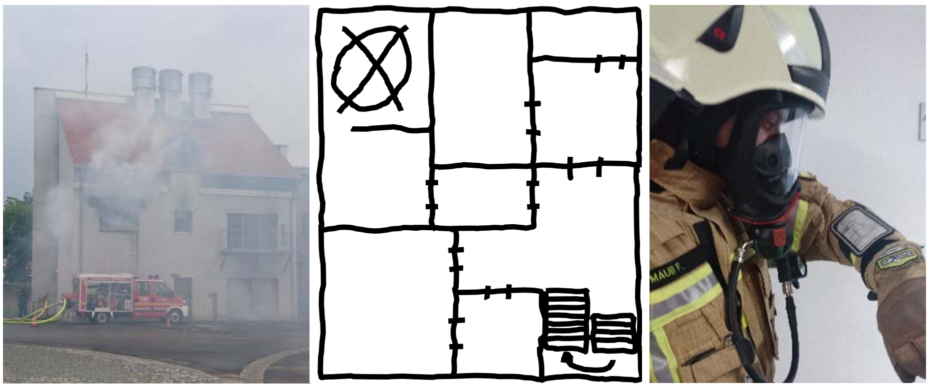


Figure 2. Fire training building (left), sketch-map (middle), and small-scale copy on the team leader's sleeve for continuous access (right)

G_2 was the second treatment group. The treatment received by teams of this group was continuous access to the before-mentioned sketch-map. Like the teams of G_1, they were briefed with the standardized phrase supported by the graphical information of the map. In addition, they also took the graphical information with them during the rescue mission. To simulate an information device, they were equipped with a small-scale copy of the sketch-map that was wrapped in a protective cover and fastened on the team leader's sleeve (cf. Figure 2). As discussed in section 5, the real-world application of our treatments – especially the continuous access to the map – will inevitably require the use of information technology. As an approximation to measure the resulting effects, however, we decided to stick with the paper-based solution.

To examine **H_1**, we compared the average task completion rates of the three groups. A task was rated as completed if the team could find the dummy and bring it outside. To examine **H_2**, we took the time the firefighters needed to complete the task. Precisely, we took the time the teams needed to find the dummy inside the apartment. The period began once the first team member crossed the door sill and ended once the first team member reached the dummy. Since there was no visual contact with the team, an instructor (who was following the team for security reasons anyway) radioed to the squad leader once the dummy had been found. By comparing the average times needed to find the dummy for the three groups (task completion time), we could test **H_2**. Finally, to examine **H_3**, we captured the task satisfaction of the participants by means of a questionnaire as described in the next subsection. Tables 1 and 2 summarize the experimental design.

Table1. Summary of the different groups

Groups
G_0: verbal only information
G_1: verbal + punctual graphical information
G_2: verbal + continuous graphical information

Table 2. Summary of the tests of Hypotheses

Test of Hypotheses
H_1: task completion rate
H_2: task completion time
H_3: After-Scenario Questionnaire (cf. 3.2)

Following the requirements of a laboratory experiment, we took several measures to ensure the internal and external validity of our results [25]. Regarding internal validity, we tried to make our treatments the only varying factor between the different groups. Consequently, we controlled the surrounding conditions as far as possible. First, all teams had to perform the identical task. The dummy had been positioned at the identical spot in the same room for all teams. The doors on the way were equally closed, and the rooms had been equally filled with smoke. All in all, the briefing at the beginning of the task was the only varying factor between the groups.

Also, the search and rescue task was chosen with the goal to ensure internal validity in mind. In a firefighting task, for example, several other factors like hose management and extinguishing technique might have affected the task completion time much more than our treatments. In our scenario, on the other hand, the teams' orientation abilities inside the apartment were the major determinant for how long it took to find the dummy. As another measure, we ensured that the participants had no prior knowledge of the apartment or the task. The teams were separated and not allowed to watch other teams during their briefings or the tasks itself. This way, we could rule out learning effects and keep the conditions equal for all teams. Finally, we captured several control variables like the firefighters' experience, age, and other in the questionnaire to rule out differences based on those factors (cf. Section 3.2).

We maximized external validity mainly by choosing the fire training building as a realistic site. Since it was specifically designed to simulate apartment fires, the task could be reproduced as realistically as possible under safe and controlled conditions. Remotely controlled gas-fires provided heat and real flames. Multiple fog machines simulated realistic smoke and limited the sight to only a few centimeters. Loudspeakers played acoustic stimuli like the crackling of the fire. All in all, the surrounding conditions of the experiment were closely comparable to those of a real-world emergency. Besides, the scenario itself was developed in roundtable sessions with instructors of the firefighting academy and characterized as a realistic everyday scenario for firefighters. Furthermore, the sample of subjects contributed to ensuring external validity. All participants were firefighters that also conduct this kind of tasks in real operations. We hence assume that our results are transferable to comparable real-life emergency operations of firefighters.

3.2 After Scenario Questionnaire

To capture feedback from our participants and assess their task satisfaction, we used a questionnaire. In its first part, we asked about general and demographic information. This way, we wanted to record several control variables that could have an influence on our results. To assess the firefighters' experience, we asked both team members for how long they have been members of a fire department and for how long they have been trained to wear SCBA. Besides that, we captured the average number of emergencies and SCBA-operations they are responding to per year. Furthermore, we asked for the command level they are normally working in and the type of fire department they are working for. Finally, the age and gender of our participants were captured. This first part of the questionnaire was filled out prior to the task.

In the second part of the questionnaire, we wanted to assess the participants' task satisfaction to examine $\mathbf{H_3}$. This part was filled out by the firefighters right after completing their task. Due to previous experiences in questioning firefighters in or shortly after stressful situations, we wanted to keep this part as short and simple as possible. Consequently, we decided to use the easy, yet well-established After-Scenario Questionnaire (ASQ) [26]. The ASQ consists of three items that we slightly adapted:

- Overall, I am satisfied with the ease of completing the task in this scenario
- Overall, I am satisfied with how long it took to complete the task in this scenario
- Overall, I am satisfied with the available supporting information when completing the task in this scenario

All three items had to be rated by the teams on a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree), as well as an n/a-point. To gather additional qualitative feedback, we included open-ended questions about what the firefighters perceived as positive and negative about the available information.

4 Results

During the experiment, we observed the performance of 150 firefighters that were composed to 75 teams. We could, however, not use all 75 observations for various reasons. For example, one team already knew the fire training building from a previous visit, another accidentally activated an emergency shutdown button, and one team was interrupted by a technical fault. All in all, we had to delete six observations from our initial sample, which resulted in an adjusted sample of 69 teams. Those remaining teams were equally distributed among the three groups, resulting in 23 observations per group.

Regarding our control variables, the participants were 27 years old on average, mainly male (>98%), have been members of a fire department for 12 years, and trained to wear SCBA for 7 years. On average, they respond to 26.48 emergency operations per year. Of those operations, they are wearing and actually applying SCBA 3.62 times a year. Looking at the command level, 24 of our participants had a team member qualification, 56 had a team leader qualification, 50 had a squad leader qualification, and the highest qualification of eight was the one of a platoon leader. Finally, most of our participants (>98%) were members of voluntary fire departments. To rule out possible influences, we calculated the correlation coefficients between our control and dependent variables. None of those coefficients are significant, however.

As stated before, our dependent variables are the task completion rate, the task completion time, and the three items of the ASQ (satisfaction with ease, satisfaction with time needed, and satisfaction with available information). We summarize the means and standard deviations of those factors for the three groups in Table 3. For all factors, there are first impressions of superiority for $\mathbf{G_1}$ and $\mathbf{G_2}$, which had graphical information available. Most dominantly, the means of the task completion times of $\mathbf{G_1}$ and $\mathbf{G_2}$ were lower than that of $\mathbf{G_0}$. Looking at the ASQ items, $\mathbf{G_1}$ showed the highest agreement, whereas $\mathbf{G_0}$ showed the lowest. Interestingly, it was not $\mathbf{G_2}$ with continuous graphical information that showed the best results, as our hypotheses implied. Instead, $\mathbf{G_1}$ with only punctual access to graphical information performed better.

Table 3. Descriptive summary and test statistics

	Group 0		*Group 1*		*Group 2*		*t-tests (p-values)*		
	mean	*SD*	*mean*	*SD*	*mean*	*SD*	*G_0 vs G_1*	*G_0 vs G_2*	*G_1 vs G_2*
Task completion rate (0 or 1)	0.87	0.34	0.96	0.21	0.87	0.34	0.15	0.50	0.15
Task completion time (sec.)	340	100	285	78	289	67	0.02*	0.03*	0.41
ASQ 1	4.30	1.74	5.22	1.59	4.61	1.62	0.04*	0.27	0.10
ASQ 2	4.43	1.38	5.17	1.40	4.57	1.50	0.04*	0.38	0.08
ASQ 3	5.52	1.73	5.96	1.58	5.83	1.56	0.19	0.27	0.39

Legend: ASQ items from 1=disagree to 7=agree; *: $p < 0.05$, one-tailed testing, N=69, each group n=23)

To examine the statistical significance of those impressions, we employed t-tests to check for differences between the groups. Since our hypotheses implied an increased task completion rate, decreased task completion time, and increased task satisfaction for $\mathbf{G_1}$ and $\mathbf{G_2}$, we used one-tailed, unpaired t-tests. The test results are summarized in Table 3. As can be seen, there were no significant differences between the groups regarding task completion rate. Consequently, $\mathbf{H_1}$, which implies that a higher degree of graphical building information would help firefighters to perform their task more accurately, cannot be accepted. We could not identify any significant differences between control group $\mathbf{G_0}$ and treatment groups $\mathbf{G_1}$ and $\mathbf{G_2}$.

For task completion time, Table 3 indicates that the teams of both treatment groups $\mathbf{G_1}$ and $\mathbf{G_2}$ were significantly faster than the teams of control group $\mathbf{G_0}$. However, a significant difference between the two treatment groups could not be observed. For a more detailed insight into the task completion times, we generated boxplots that are shown in Figure 3. The more robust measures of the boxplots confirm the beforementioned results. The medians and quartiles of $\mathbf{G_1}$ and $\mathbf{G_2}$ are consistently below the ones of $\mathbf{G_0}$. The only visible difference between the treatment groups is the slightly higher median of $\mathbf{G_2}$. Overall, $\mathbf{H_2}$, which implies that firefighters provided with a higher degree of graphical building information would need less time to complete their task, can be partially accepted. Both forms of graphical information did make a significant difference compared to the verbal instruction. The degree of graphical information (i.e., punctual or continuous access) showed no further impact.

Looking at the task satisfaction, there are two significant differences in the ASQ item values (cf. Table 3). Members of $\mathbf{G_1}$ were more satisfied with the ease of completing the task and the time needed to complete it than members of $\mathbf{G_0}$. On the other hand, there were no significant differences regarding the satisfaction with the available information. In addition, the values of $\mathbf{G_2}$ show no significant difference to the other groups at all. Consequently, $\mathbf{H_3}$, which implies that a higher degree of graphical building information increases firefighters' task satisfaction, can again be partially accepted. The punctual access to graphical information during the briefing did raise the task satisfaction, while the continuous access to graphical information during the mission did not.

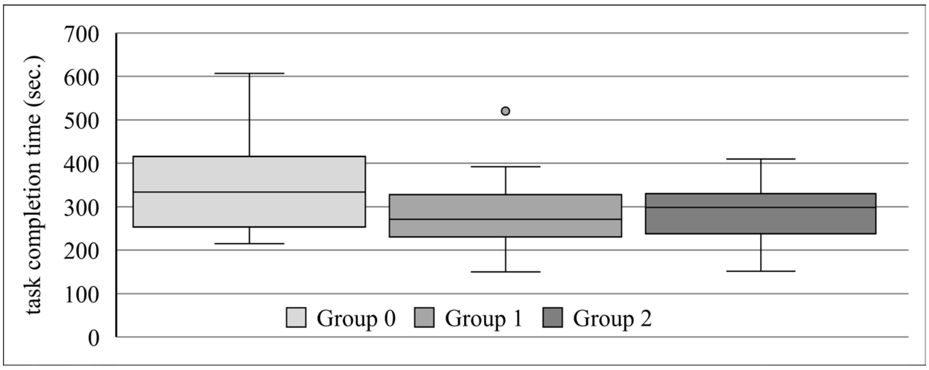


Figure 3. Boxplots for task completion time by groups

5 Discussion

In this section, we discuss our results in more detail. The discussion is informed by qualitative feedback gathered during the experiment, insights from reviewing the results with instructors of the firefighting academy, and scientific literature.

As shown in the previous section, our results only partially support the hypotheses. Regarding the task completion rate, there were no significant differences between the groups at all. Looking at the firefighters' qualitative feedback, all teams that were not able to complete their task stated physical exhaustion as the reason. They all managed to find the dummy but had to give up on their way out. None of the tasks failed due to insufficient information or lost orientation. Consequently, it seems that there has been no impact of our treatments toward the task completion rate.

With punctual graphical information, however, the firefighters could perform their task faster and were more satisfied than with verbal information. This outcome also corresponds to qualitative feedback from the participants. In $\mathbf{G_0}$, they criticized that they received *"too much information in a short time"* and that the information was *"misleading."* In $\mathbf{G_1}$, on the other hand, they praised the *"short overview"* that was *"well understandable"* and allowed a *"better orientation."* Opposed to our hypotheses, the continuous access to the graphical information did not bring further improvements. The firefighters belonging to $\mathbf{G_2}$ could perform their task faster than those of $\mathbf{G_0}$. The task satisfaction, however, was not significantly different. Between $\mathbf{G_1}$ and $\mathbf{G_2}$, there were no differences, at all. Possible reasons for this can again be found in the qualitative feedback. In $\mathbf{G_2}$, the firefighters saw a too strong *"distraction"* and *"fixation"* to the map that led to the problem of *"blindly relying on it"* and *"overlooking things."* Consulting the instructors of the firefighting academy, we identified the low complexity of the building as an additional explanation why there were no advantages in the continuous access to the graphical information. The layout of the apartment and the way to the dummy as it is displayed in Figure 2 might have been simple enough that the firefighters did not experience cognitive overload [22]. According to the instructors, though, a real-world apartment fire is rarely more complicated. It seems that for the scenario of apartment fires, a short look on the map simply suffices. In larger events with more complicated layouts like multiple burning apartments or a burning hospital, however, they see potential in the continuous access to the information.

Looking at the practical adoption, our experiment did not include the creation of the sketch-map. Since we wanted to provide the teams with standardized briefings for internal validity, the squad leader position was taken by a member of the research team. In reality, the squad leader would have to ask the house owner or neighbors for information about the building and draw the map according to it. The time effort of drawing must, of course, be opposed to the time gained by completing the task faster. According to the instructors, however, the squad leader would question knowledgeable attendants anyway. He would also have some short time to draw a map while the team is grabbing their equipment and preparing to enter the building. Of course, the much more favorable solution would be that the graphical building information is available already in advance to an emergency. This would save time and is not depending on the availability of attendants. One solution would be that house owners provide their local fire department with building plans and the department stores them in a database.

As stated before, the application in practice might benefit from using information technology. Of course, the first treatment of punctually presenting the graphical information may also work with simpler means like pen and paper. It is, however, unclear if the squad leader could be able to draw a sketch-map faster with electronic means like a drawing software on a tablet computer which provides predefined design elements. This should be examined in further experiments. In the previously discussed case that the graphical information might be available prior to an emergency, the pure amount of data would call for the use of information technology. Corresponding FITs that are discussed in literature are digital plans [1, 2] and on-site emergency response information systems [5, 6]. Making the information available continuously will, however, require the use of additional, more intricate information technologies. On the one hand, the information would have to be duplicated in real-time. While the sketch-map is given to the team, another instance of it should remain with the squad leader for his commanding duties on site. On the other hand, the map taken by the team would have to be heat-resistant, waterproof and readable in dark smoke. Instead of the paper approximation that we used in our experiment (cf. Figure 2), different FITs might be used in practice. Displays could be integrated into the sleeves of intelligent protective clothing [9]. Also, augmented reality systems could be used to display the graphical information in the firefighters' breathing masks [7]. A third approach was proposed by one of the participants. For the search for fires and victims, firefighters often carry infrared cameras with them. The displays of those cameras could also be used to show the sketch-map. This way, the firefighters would not have to carry any additional devices with them but could use systems they are already used to. It might indeed be the simplest, yet most acceptable solution for the practitioners. The performance of all those technological solutions should be examined in future research.

All in all, our results hold several implications for both academia and practice. Regarding academia, we present a task-centered approach to identify information potentials for a specific scenario and propose suitable technological solutions for firefighters. This constitutes an alternative approach compared to the mostly technology-driven development of FITs. The overall procedure can be transferred to many other problems. First, other scenarios from the firefighter domain can be examined in a similar way. From salvage rescues after car accidents to forest fires, there are many possible scenarios. Besides, the procedure can be applied to scenarios of other first responders like police or rescue departments as well. In the work at hand, we build on the theory of situation awareness. While this theory is well established in emergency management, it has by now mainly been used to examine higher levels of severity. In our study, we demonstrate its applicability in a comparably small-scale and everyday scenario. To increase situation awareness, we furthermore combine this theory with findings from cognitive science. In so doing, we also show the applicability of those findings in the special domain of firefighting and could even partly confirm them.

Regarding implications for practice, we showed a potential adaption of the firefighters' conventional procedure of getting to work as fast as possible, even with insufficient information. In our experiment, we could demonstrate the positive effects of improved situation awareness. By providing the firefighters with suitable and easy to process information like a sketch-map, the task completion time could be significantly reduced. Consequently, it might make sense for the squad leaders to take a little more time to establish situation awareness in order to gain overall time advantages. The involved instructors of the firefighting academy also support the findings of the experiment. They meanwhile integrate those findings into the squad leader courses held at the academy. They try to sensitize the firefighters for the importance of situation awareness even in time-critical operations and discuss how it can be established. All in all, our study demonstrates the potential of specific information for the firefighters. The developers of FITs can use our findings to provide firefighters with such information.

Although we implemented several measures to ensure internal and external validity, there are some limitations. First, we concentrated on the specific scenario of an apartment fire. Accordingly, our results will only be transferable to real operations of this type. By now, we only used paper approximations to display the information. Employing information technologies might increase or decrease the identified effects. Therefore, as a point of future research, the graphical information should be implemented in according technologies and analyzed in additional experiments. As stated above, the scenario was realistic but not very complex. To examine the potential of continuous access to the information in detail, more complex scenarios than search and rescue during an apartment fire should be analyzed. Furthermore, 98% of our subjects were members of a voluntary fire department. While voluntary and professional firefighters perform equal tasks in Germany, there might still be differences regarding the fitness level or training frequency. To strengthen our results, an additional experiment with professional firefighters might be wise. Finally, our sample of subjects consisted of firefighters from all over Bavaria. Since firefighting tasks and training are comparable in Germany, our results can be transferred to German firefighters. Due to differences in the organization of fire departments and emergency response processes, however, we cannot straightforwardly transfer the results to other countries.

6 Conclusion and Outlook

In an attempt to facilitate the design and implementation of suitable information technologies that provide adequate support for firefighters during critical everyday missions, we presented the results of a controlled experiment, in which we examined the efficacy of different kinds of information during the search and rescue task. During the planning of our study, the search and rescue task was identified as an important task, in which the performance might be affected negatively due to an insufficient information base. At the same time, various FITs have been suggested to also facilitate this task. In our study, we found indications that certain kinds of information about the building and the victims might be effective in enhancing the task performance. However, our findings also provide indications that the continuous access to such information (as it is, for instance provided with augmented reality devices) might not provide a further improvement. Judging from the findings of our study, the design of FITs hence ought to be carefully aligned to the task that is to be supported and the information demands.

In future iterations, we plan to strengthen both the theoretical implications and the conceptual foundations of our research. To examine the robustness of our findings, we will further increase the sample size. Moreover, we plan to re-evaluate our results in different scenarios and settings, for instance with varying building and situation complexity, with members of other types of departments (such as professional firefighters), or in settings where the squad leader only has ambiguous or incomplete information. Based on the gathered results, we plan to implement and test FITs that provide theoretically grounded and empirically proven support for the search and rescue task. To better guide the implementation of such FITs, we also intend to connect our findings to existing literature about indoor-navigation, wearable IT, and other domains, which might deliver additional insights for the task-centered design of FITs.

Although we have not formally tested it, we believe that the proposed procedure to examine the existing information demand and to explore ways to provide an adequate information basis before designing and implementing FITs based on their technological capabilities can also be used to support other tasks during the fire emergency response process. In future iterations, we therefore also intend to deduce a generally usable methodology to design FITs based on relevant theories and empirical evidence. With the presented study, we hope to provide a starting point to change the development of FITs from a primarily technology-driven into a mainly demand-driven process.

References

1. Johnson, R.: A Case Study in Multiagency GIS for Managing a Large-Scale Natural Disaster. In: van Oosterom, P., Zlatanova, S., Fendel, E.M. (eds.) Geo-information for Disaster Management, pp. 155-170. Springer Berlin, Heidelberg, Berlin, Heidelberg (2005)
2. Takahagi, K., Ishida, T., Noda, S., Sakuraba, A., Uchida, N., Shibata, Y.: Proposal of the Fire Fighting Support System for the Volunteer Fire Company. In: 29th International Conference on Advanced Information Networking and Applications Workshops, pp. 484-489. IEEE Computer Society, Washington, DC (2015)
3. van Persie, M., Oostdijk, A., Fix, J., van Sijl, M.C., Edgardh, L.: Real-time UAV based geospatial video integrated into the fire brigades crisis management GIS system. International archives of the photogrammetry, remote sensing and spatial information sciences 38, 173-175 (2011)
4. Barrado, C., Messeguer, R., Lopez, J., Pastor, E., Santamaria, E., Royo, P.: Wildfire monitoring using a mixed air-ground mobile network. IEEE Pervasive Computing 9, 24-32 (2010)
5. Yang, L., Prasanna, R., King, M.: On-Site Information Systems Design for Emergency First Responders. Journal of Information Technology Theory and Application 10, 5-27 (2009)
6. Ha, K.M.: In pursuit of ideal emergency planning in Korea. Disaster Prevention and Management 21, 624-633 (2012)
7. Klann, M., Geissler, M.: Experience Prototyping: A New Approach to Designing Firefighter Navigation Support. IEEE Pervasive Computing 11, 68-77 (2012)
8. Bretschneider, N., Brattke, S., Rein, K.: Head Mounted Displays for Fire Fighters. In: 3rd International Forum on Applied Wearable Computing. VDE Verlag, Berlin (2006)
9. http://www.smartatfire.eu/media/37543/Rossou.pdf

10. Salim, F., Belbasis, A., Prohasky, D., Houshyar, S., Fuss, F.K.: Design and evaluation of smart wearable undergarment for monitoring physiological extremes in firefighting. In: ACM International Symposium on Wearable Computers: Adjunct Program, pp. 249-254. ACM, New York, NY (2014)
11. Schlauderer, S., Overhage, S., Weidinger, J.: New Vistas for Firefighter Information Systems? Towards a Systematic Evaluation of Emerging Technologies from a Task-Technology Fit Perspective. In: 49th Hawaii International Conference on System Sciences, pp. 178-187. IEEE Computer Society, Washington, DC (2016)
12. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37, 32-64 (1995)
13. Endsley, M.R.: Theoretical Underpinnings of Situation Awareness: A Critical Review. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness Analysis and Measurement. Lawrence Erlbaum Associates, Mahwah, NJ (2000)
14. Pagotto, J., O'Donnell, D.: Canada's multi-agency situational awareness system – Keeping it simple. In: 9th International Conference on Information Systems for Crisis Response and Management. Simon Fraser University, Vancouver, Canada (2012)
15. Ley, B., Pipek, V., Reuter, C., Wiedenhoefer, T.: Supporting Inter-Organizational Situation Assessment in Crisis Management. In: 9th International Conference on Information Systems for Crisis Response and Management. Simon Fraser University, Vancouver, Canada (2012)
16. Kaplan, S., Kaplan, R., Sampson, J.R.: Encoding and arousal factors in free recall of verbal and visual material. Psychonomic Science 12, 73-74 (1968)
17. Paivio, A., Rogers, T.B., Smythe, P.C.: Why are pictures easier to recall than words? Psychonomic Science 11, 137-138 (1968)
18. Sampson, J.R.: Free recall of verbal and non-verbal stimuli. Quarterly Journal of Experimental Psychology 22, 215-221 (1970)
19. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. Cognitive Science 11, 65-100 (1987)
20. Mousavi, S.Y., Low, R., Sweller, J.: Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. Journal of Educational Psychology 87, 319-334 (1995)
21. Mayer, R.E., Sims, V.K.: For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning. Journal of Educational Psychology 86, 389-401 (1994)
22. Mayer, R.E., Roxana, M.: Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist 38, 43-52 (2003)
23. Frøkjær, E., Hertzum, M., Hornbæk, K.: Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: SIGCHI Conference on Human Factors in Computing Systems, pp. 345-352. ACM, New York, NY (2000)
24. ISO/IEC: Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs) - Part 11: Guidance on Usability. International Organization for Standardization (1998)
25. Cox, D.R., Reid, N.: The Theory of the Design of Experiments. Chapman & Hall/CRC, New York, NY (2000)
26. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. SIGCHI Bulletin 23, 78-81 (1991)

Track 10:

Human-Computer Interaction

Track Chairs:

Prof. Dr. Rene Riedl
Johannes-Kepler-Universität Linz

Prof. Dr. Alexander Mädche
Karlsruher Institut für Technologie (KIT)

Dr. Mario Nadj
Karlsruher Institut für Technologie (KIT)

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Towards a Taxonomy of Platforms for Conversational Agent Design

Stephan Diederich[1], Alfred Benedikt Brendel[1], Lutz M. Kolbe[1]

[1] University of Goettingen, Chair of Information Management, Goettingen, Germany
stephan.diederich@stud.uni-goettingen.de,
{abrendel, lkolbe}@uni-goettingen.de

Abstract. Software that interacts with its users through natural language, so-called conversational agents (CAs), is permeating our lives with improving capabilities driven by advances in machine learning and natural language processing. For organizations, CAs have the potential to innovate and automate a variety of tasks and processes, for example in customer service or marketing and sales, yet successful design remains a major challenge. Over the last few years, a variety of platforms that offer different approaches and functionality for designing CAs have emerged. In this paper, we analyze 51 CA platforms to develop a taxonomy and empirically identify archetypes of platforms by means of a cluster analysis. Based on our analysis, we propose an extended taxonomy with eleven dimensions and three archetypes that contribute to existing work on CA design and can guide practitioners in the design of CA for their organizations.

Keywords: Conversational agent, chatbot, design science, taxonomy, cluster analysis

1 Introduction

As artificial intelligence, particularly machine learning, increasingly permeates and impacts our daily private and professional lives, it drives a new wave of technological change and unprecedented automation of cognitive tasks [1]. One phenomenon in this wave are continuously improving conversational agents (CAs) which benefit from expanding functionalities and the diffusion of powerful and connected (mobile) devices. The presence of CAs is more and more increasing, such as in the form of Apple's Siri, Amazon's Alexa or in-car assistants. Basic CAs conduct information search for us, send messages or enter meetings in a calendar. Similarly, more and more companies use CAs for different purposes, such as automation and innovation in customer service or marketing and sales [2–7]. CAs can be distinguished from other software by their ability to interact with users based on natural language. This language can be spoken, as for example in the case of Amazon's Alexa, or written, often referred to as chatbots. In recent years, CA capabilities significantly expanded from simple rule-based systems to seemingly intelligent assistants [5, 8, 9] as a result of advances in machine learning and natural language processing.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

In research, CAs attracted increasing interest in the last few years with different foci, such as information disclosure of users [10, 11], human performance improvement [12] or user authenticity perception [8]. In parallel with increased research interest in the IS community, organizations have started to experiment with and introduce CAs, often in the context of larger artificial intelligence initiatives [4, 8, 13]. However, many CAs fell behind expectations and often disappeared due to flaws related to their design, thus successful design remains a complex challenge in practice where various aspects need to be addressed [5, 14, 15].

With the popularity of CAs in both research and practice, a variety of enterprise CA platforms has emerged, supporting the design of CA with different functionality [16]. This includes both offerings of established technological players, such as Google's DialogFlow, as well as start-ups specialized in CAs such as ManyChat. While several studies can inform CA design through principles of form and function [5, 17, 18], the platforms that are used to actually designing CAs, providing both possibilities as well as constraints for the implementation, have not been studied in the past to the best of our knowledge. In order to gain a better understanding of these novel platforms, we first study along which dimensions CA platforms can be categorized (RQ1). Building on these dimensions and empirical data, we then aim to identify archetypes of platforms and their distinctive characteristics (RQ2). To address these research questions, we first develop a taxonomy of CA platforms, both conceptually from a literature review and empirically through the iterative classification of platforms. We then perform a cluster analysis to identify archetypes and gain a better understanding of commonalities and differences between the platforms.

We continue by describing the research background on CAs and presenting our research approach, i.e. taxonomy development followed by a cluster analysis. Finally, we present and discuss our results, particularly the developed taxonomy and identified archetypes, and close by suggesting directions for future work on CAs.

2 Research Background

The basic idea of a CA is to interact with users using natural language just like in a human-to-human conversation [19] and exchange information through verbal communication about a common topic [20]. This idea dates back decades to the 1960s when the first CA, called ELIZA, was developed by Joseph Weizenbaum [21]. Since then, a variety of CAs emerged (and often disappeared) that used simple pattern matching to provide a set of responses to the users [5, 22]. With recent technological advances, particularly in the fields of machine learning and natural language processing, as well as the diffusion of powerful, connected devices, the capabilities and potential of CAs increased significantly and they moved from rule-based systems to seemingly intelligent agents [22, 23]. Due to this development, CAs regained momentum in research and practice in the past few years and a variety of new CA offerings emerged.

In order to organize this variety that is available today, Gnewuch et al. [5] provide a simple taxonomy that consists of two dimensions including primary mode of communication and context (see Table 1). As natural language can be written or spoken [24], the mode of communication indicates the primary way in which users interact with a CA. For example, Apple's virtual assistant Siri is accessed using voice commands whereas Spotify's messenger bot works using digital text messages. CAs with text-based input are often referred to as chatbots in research as well as practice [2, 25, 26], while CAs with speech-based input are described as virtual or digital assistants [25, 27]. Because voice input can be quite easily transferred to written input in most cases, the boundaries between the mode of communication are often blurred as bots offer both spoken and written language as input. For example, a customer can request a ride with Lyft both via chat, e.g. Facebook Messenger or Slack, and by voice command, for example with Amazon Echo [28].

The second dimension, context, indicates whether the CA serves a specific domain such as a task or business function, or can interact on any topic with its users [5, 29]. General-purpose CAs like text-based Cleverbot [30] and Mitsuku [31] can have a conversation about any topic and continuously learn as they interact with users. For speech-based, general-purpose CAs the most prominent examples are from private life, such as Siri or Google Assistant.

Table 1. Classification of CA according to Gnewuch et al. [5]

		Context	
		General-purpose	**Domain-specific**
Communi-cation mode	**Text-based***	ELIZA, Cleverbot, Chatterbot, Mitsuku, …	Enterprise-class CAs, IKEA's Anna, Starbucks Chatbot, …
	Speech-based**	Apple's Siri, Amazon's Alexa, Google Assistant, …	SPECIES [29], in-car assistants, speech-based service agents, …

*Text-based: Chatbot, chatterbot, dialogue system, etc.
**Speech-based: (Virtual) personal assistant, digital companion, smart agent, etc.

Domain-specific CAs include a wide variety of CAs, for example in a professional context for internal and external purposes, such as customer service [4, 8], IT service desk tasks, product marketing [3], and e-commerce [14]. Further exemplary domains from private life include museums [32, 33] and healthcare [34].

In order to design a CA, a variety of development platforms exists to model a bot's behavior and to deploy them, for example on Facebook or by embedding the CA in the company website. Such platforms are characterized by an extensible technological foundation, i.e. the natural language processing and machine learning capabilities, created by a platform owner, on top of which developer can build platform-augmenting applications [35], such as conversational agents for a specific domain and organization.

The development platforms offer different ranges of functionality regarding aspects such as the bot's implementation, continuous training, analytics or hosting. With regard to the implementation for example, the platform Chatfuel [36] offers to quickly model a bot's behavior within a few minutes using a web interface while Twyla [37] uses supervised learning to automatically learn from existing data, such as customer service conversations or product catalogues. Concerning analytics, the functionality of CA platforms ranges from basic analysis (e.g. number and length of conversations) to advanced approaches, such as automatic sentiment and topic detection. Overall, a large number of enterprise platforms exists that allows building and introducing both text- and speech-based CA for general-purpose or specific domains.

3 Research Approach

In order to determine the distinct characteristics of CA platforms (RQ1) and to empirically identify archetypes (RQ2), we develop a taxonomy and perform a cluster analysis after classifying the respective platforms. The role of taxonomies is well recognized in information systems (IS) as they provide structure and organize knowledge in a field [38–41]. Within IS research, a multitude of taxonomies has been developed, covering for example business models of FinTechs [42], (mobile) health IT [43, 44] or cybercrime [45]. In particular in a diverse, emerging research area, taxonomies can provide useful insights into the grouping of objects based on their common characteristics [41].

To create our taxonomy, we follow the method proposed by Nickerson et al. [41] which iteratively develops a taxonomy based both on existing conceptual knowledge as well as empirical observation. This method clearly defines the necessary steps and ending conditions, providing a rigorous and useful approach for the systematic creation of a taxonomy, and to avoid the risk of defining and altering dimensions and characteristics through ad-hoc changes. The Nickerson method has been successfully applied to develop a variety of taxonomies, such as for collaborative applications [46] or carsharing business models [47]. Our complete research approach consists of three phases and is summarized in Table 2.

Table 2. Research approach phases

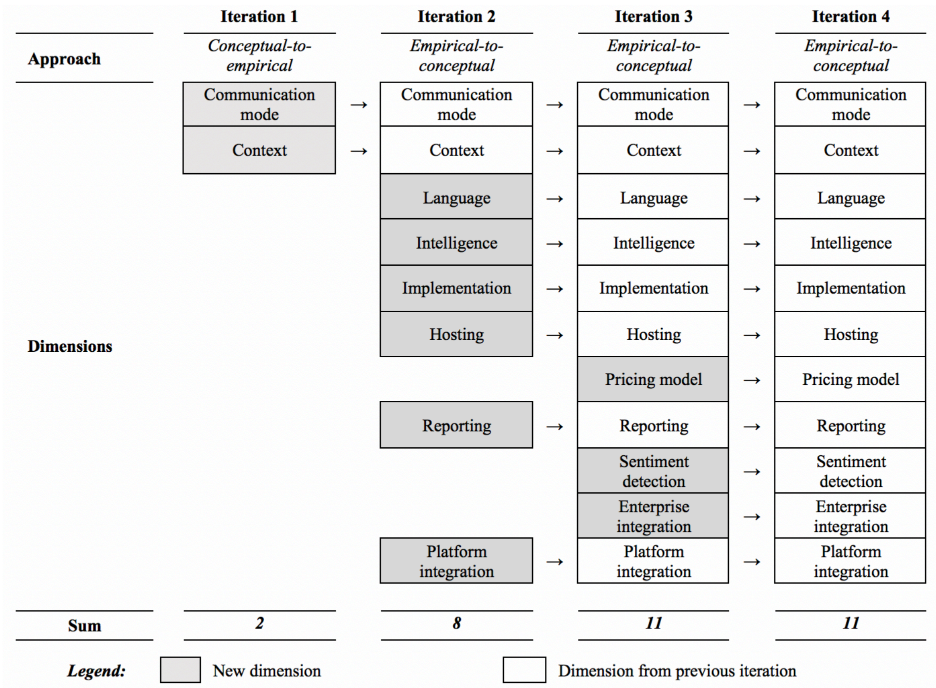
	Phase 1: Create database	**Phase 2: Develop taxonomy**	**Phase 3: Conduct cluster analysis**
Steps	• Search for CA platforms in CrunchBase and on the web • Request additional information where required	• Define meta-characteristic for the taxonomy • Iterate through taxonomy development until ending conditions are met	• Determine useful number of clusters • Specify the companies belonging to each cluster
Method	Lit. review, desk research	Taxonomy development	Clustering algorithms
Source	CA lit., blogs, practice reports, CrunchBase	CA literature, CA platform database	Taxonomy of CA platforms with empirical data
Results	Database with 51 CA platforms	Taxonomy of CA platforms with 11 dimensions	Three identified CA platform archetypes

Phase 1: Set up database: The first research phase aimed at the creation of a database with CA platforms that were operational in May 2018. For this we examined existing literature on CA, searched the world's largest startup database (CrunchBase), a variety of blogs (e.g. https://chatbotsjournal.com), and industry reports (e.g. Oracle [13]). For our search, we used the terms "conversational agent" with the synonyms "chatbot" and "digital assistant" in combination with "design" and "platform". Platforms that were not operational (i.e. actively providing the option to create a CA) were excluded from the database. Missing or incomplete data, particularly on pricing models, was gathered via e-mail requests. At the end of the first research phase, we created a database with 51 platforms for CA design.

Phase 2: Develop taxonomy: The objective of the second phase was to create a taxonomy of CA platforms that contains the most important dimensions along which the platforms differ based on the method described by Nickerson et al. [41]. For our research, we defined CA development platforms as the meta-characteristic for the taxonomy from which all subsequent dimensions follow. Regarding the ending conditions that indicate whether the taxonomy development process is completed, we used the eight objective (such as all objects have been examined and no new dimension or characteristics were added in the last iteration) and five subjective ending conditions (concise, robust, comprehensive, extendible and explanatory) from Nickerson et al. [41]. We started the taxonomy development with a conceptual-to-empirical iteration. In this initial iteration we added two dimensions (CA primary mode of communication, CA context [5, 29]) that were identified in our literature review (see Table 1). The following three iterations were empirical-to-conceptual and added nine dimensions in total, such as pricing model, implementation mode or hosting (see Figure 1). After all platforms in our database were successfully classified and both subjective and objective ending conditions were met, we considered the taxonomy final.

Phase 3: Perform cluster analysis: The objective of the third research phase was the empirical identification of CA platform archetypes (RQ2). For this purpose, we conducted a cluster analysis. Cluster analysis aims at grouping objects where objects in one group are as similar as possible and as dissimilar as possible from objects in other groups [48]. Following the recommendations by Punj and Stewart [49] to first determine the number of clusters and subsequently use an iterative partitioning technique like k-means, we chose a two-stage clustering approach: First, we defined the number of clusters with Ward's method. With this method we agglomeratively clustered (i.e. repeatedly combined the two closest objects into one group until all objects belong to the same group [50]) the CA platforms using SPSS version 25 and squared Euclidean distance. We then reviewed the descriptive data on these iterations, i.e. the coefficient distance, the dendrogram and the scree plot using the elbow rule. These indicated that a three cluster would be most useful. In the second step, we used the chosen number of groups for a k-means clustering procedure. The procedure used three iterations until no significant enhancements were achieved.

Figure 1. Iterations of our taxonomy development



4 Results

In the following, we present our taxonomy for CA platforms (RQ1) and provide examples for platforms to demonstrate their respective characteristics. We then continue with describing the archetypes of platforms we identified in the two-step cluster analysis (RQ2).

4.1 Taxonomy for CA platforms

The resulting taxonomy consists of 11 dimensions with two to four characteristics each (see Table 3). The first two dimensions were found in existing literature [5]. Each platform was assigned one characteristic for each dimension. We omitted dimensions that were the same across all platforms (representation of the CA with an avatar, assigning a name to the CA) as we aim to distinguish them by their main characteristics.

The first dimension, *Communication mode*, refers to the primary way with which a user communicates with a CA and may more broadly described as the user interface, i.e. text-based, speech-based or both [5]. For example, platforms such as ManyChat, pandorabots, or Recime exclusively offer building text-based CAs (referred to as Chatbots) whereas aivo and The Pullstring Platform focus on agents that interact with its users via speech. Furthermore, platforms such as Nuance and IPSoft offer to build and integrate CA that interact via both text and speech. The dimension *context* indicates in which task or business domain a CA built on the respective platform can be used [5]. For example, SurveyBot offers to build specific CAs that conduct surveys and collect their results or Octane AI's CA that provides sales optimization by interactively engaging with users that abandon their digital shopping carts. The dimension *language* refers to the language(s) supported by the CA where platforms offer support for single languages (mostly English, e.g. botmother) or multiple languages (e.g. ChatClub). *Intelligence* indicates whether a CA is primarily based on rules that perform rather simple pattern matching, such as ChatbotsBuilder, or has the ability to self-learn, such as Twyla, enabling the CA to improve over time as it converses with its users.

Table 3. Taxonomy of CA platforms

<table>
<tr><th>Dimension</th><th colspan="12">Characteristics</th></tr>
<tr><td>Communication mode</td><td colspan="4">Text-based</td><td colspan="4">Speech-based</td><td colspan="4">Both</td></tr>
<tr><td>Context</td><td colspan="6">General-purpose</td><td colspan="6">Domain-specific</td></tr>
<tr><td>Language</td><td colspan="6">Single language</td><td colspan="6">Multi language</td></tr>
<tr><td>Intelligence</td><td colspan="6">Rule-based</td><td colspan="6">Self-learning</td></tr>
<tr><td>Implementation</td><td colspan="3">Programming</td><td colspan="3">Modeling</td><td colspan="3">Supervised learn.</td><td colspan="3">Hybrid</td></tr>
<tr><td>Hosting</td><td colspan="4">On-premise</td><td colspan="4">Cloud</td><td colspan="4">Both</td></tr>
<tr><td>Pricing model</td><td colspan="3">Usage-based</td><td colspan="3">User-based</td><td colspan="3">Instance-based</td><td colspan="3">Free</td></tr>
<tr><td>Reporting</td><td colspan="6">Without reporting</td><td colspan="6">With reporting</td></tr>
<tr><td>Sentiment detection</td><td colspan="6">Without sentiment</td><td colspan="6">With sentiment</td></tr>
<tr><td>Enterprise integration</td><td colspan="4">None</td><td colspan="4">API</td><td colspan="4">Pre-build interface(s)</td></tr>
<tr><td>Platform integration</td><td colspan="6">Single-platform</td><td colspan="6">Cross-platform</td></tr>
</table>

Existing dimension

New dimension

The dimension *Implementation* indicates how a bot is built, whether via programming (actually writing code), modeling (modeling typical user conversations in a flow chart), supervised learning (training the CA with labeled conversations), or with the help of a hybrid approach (e.g. modeling in combination with supervised learning). Popular platforms for creating a bot via programming are wit.ai, and Zenbot. With regard to modeling, the most common platforms used to build bots include Massively, ManyChat, and LeadFlip. In contrast to programming and modeling, some platforms such as Twyla rely on training a CA with existing user interactions (supervised learning) while others like Creative Virtual and gupshup use a combination of these implementation approaches. *Hosting* refers to the deployment of CAs where platform offerings range from on-premise (e.g. botpress), public cloud (e.g. ChatterOn or Converse), and both methods combined. *Pricing* refers to the pricing model that is used by the platform. The models we observe in our data include usage-based (i.e. based on number of interactions, such as Microsoft Azure Bot), user-based (i.e. based on number users, such as MobileMonkey), instance-based (i.e. based on number of CA, such as ChatbotsBuilder) and free (such as It's Alive).

Reporting indicates whether a CA platform offers reporting functionality to monitor the CA's interactions and usage, such as number of conversations or unique users (for example provided by reply.ai and Lex). *Sentiment detection* indicates whether a platform allows automatic detection of user sentiment during an interaction. Finally, *Enterprise integration* indicates whether a CA platform offers pre-built interfaces or APIs to let CAs access different enterprise systems such as a CRM for information that is used in a conversation with a user. For example, Microsoft Azure Bot Service can automatically retrieve information from its Dynamics CRM in a user interaction via a standardized interface. Other platforms, for example pandorabots or Rasa, can retrieve data from enterprise systems via API calls.

4.2 Archetypes of CA platforms

The three clusters contain 18 (cluster 1), 19 (cluster 2), and 14 (cluster 3) platforms from our database (Table 4). Each cluster has different centers along the dimensions of the taxonomy developed in this study. As the characteristics within the taxonomy are mutually exclusive and collectively exhaustive, we describe the clusters with a crosstab analysis showing percentages for each characteristic within a cluster (see Figure 2). For example, 22% of all CA platforms in cluster 1 support a single language whereas 78% offer multi language support. In the following, we describe the clusters, highlight their distinctive characteristics, and provide illustrative examples.

Archetype 1 – Multi-language, integrative CA platform with advanced analytical functionality: The first cluster contains platforms that mainly support multiple languages, self-learn over time, and integrate with different enterprise systems, such as CRM software, as well as various platforms, such as social media. All platforms within this cluster offer reporting functionalities and the majority of platforms has built-in sentiment detection. These platforms include the CA offerings of major technology players, such as Oracle Intelligent Bots, Microsoft Azure Bot Service, IBM Watson Assistant or Amazon Lex, and large technology companies that strive to automate tasks particularly in customer service, IT operations as well as product and marketing like IPSoft or Nuance. CA platforms in this cluster support text-based and speech-based communication and include CAs for various purposes. Whereas platforms in cluster 2 and 3 mainly focus on the modeling of typical conversation flows as an implementation approach, platforms in this clusters also offer supervised learning (allowing to train a CA with a set of historical, labelled data) and hybrid approaches (i.e. a combination of modeling and supervised learning). Regarding deployment, many platforms offer cloud or cloud and on-premise hosting and pricing depends on actual usage.

Archetype 2 – General-purpose, cloud-based CA platform with single language and API support: The second cluster includes platforms that focus on CAs for different purposes, support a single language (in most cases English), and are primarily hosted in the cloud. With regard to integration with other enterprise software, these platforms typically offer APIs to program the automatic retrieval of data from existing systems, such as CRM. Examples of platforms in this cluster include pandorabots, Recime and Xenioo. These platforms mostly use modeling as the implementation approach, as in the first cluster. Regarding the analytical functionality, none of the platforms provide sentiment detection, while about two third of the platforms in this cluster offer reporting features. Regarding the integration of CAs with target platforms, the companies within this cluster are split between single-platform (e.g. TalkBot for Facebook) and cross-platform support (e.g. pandorabots).

Figure 2. Cross tab analysis

Dimension	Characteristics	Archetype 1	Archetype 2	Archetype 3
	Number of platforms in cluster	*18*	*19*	*14*
Communication mode	Text-based	33%	68%	100%
	Speech-based	33%	5%	0%
	Both	33%	26%	0%
Context	General-purpose	83%	100%	21%
	Domain-specific	17%	0%	79%
Language	Single language	22%	89%	93%
	Multi language	78%	11%	7%
Intelligence	Rule-based	0%	0%	50%
	Self-learning	100%	100%	50%
Implementation	Programming	6%	11%	0%
	Modeling	50%	84%	100%
	Supervised learning	33%	5%	0%
	Hybrid	11%	0%	0%
Hosting	On-premise	11%	0%	7%
	Cloud	39%	89%	93%
	Both hosting	50%	11%	0%
Pricing model	Usage-based	89%	79%	50%
	User-based	6%	5%	14%
	Instance-based	0%	5%	14%
	Free	6%	11%	21%
Reporting	Without reporting	0%	26%	50%
	With reporting	100%	74%	50%
Sentiment detection	Without sentiment	39%	100%	93%
	With sentiment	61%	0%	7%
Enterprise integration	None	0%	0%	71%
	API	11%	95%	29%
	Pre-build interface(s)	89%	5%	0%
Platform integration	Single-platform	0%	47%	79%
	Cross-platform	100%	53%	21%

Archetype 3 – Text-based, domain-specific CA platform with modeling functionality: The third and final cluster contains platforms that show different distinctive characteristics: First, these platforms exclusively offer text-based CAs, which tend to be chatbots that are used in specific domains and mostly on single platforms. For example, SurveyBot can conduct interactive surveys and collect their results via Facebook Messenger. CA platforms in this cluster typically host their CA in their own clouds and pricing is based on actual usage. With regard to the capability for integration of data from other enterprise software, the majority of platforms in this cluster does not offer an API or pre-built interfaces connecting the CA to existing systems.

Table 4: Clustered platforms

Archetype 1	Archetype 2	Archetype 3
[24]7 AI	BotEngine	ChatbotsBuilder
aivo	botmother	ChatClub
BotCore	Botsify	ChatterOn
botpress	Chatfuel	E.D.D.I.
Creative Virtual	Conversation one	HubSpot / motion.ai
Dialogflow	Converse	It's alive
gupshup	Flow xo	LeadFlip
IBM Watson Assistant	Landbot.io	ManyChat
inbenta	pandorabots	Massively
Interactions	Parlo	MobileMonkey
IPSoft	Rasa	Octane AI
Lex	Recime	rebot.me
Microsoft Azure Bot Service	Sequel	Surveybot
Next IT	Smooch	Zelp
Nuance	TalkBot	
Oracle Intelligent Bots	The PullString Platform	
reply.ai	Wit.ai	
Twyla	Xenioo	
	Zenbot	

5 Discussion

In the following, we discuss the developed taxonomy and identified archetypes against the background of existing research, followed by a description of limitations of this study, and an overview of opportunities for future research.

5.1 Taxonomy and Archetypes

The taxonomy and archetypes from our analysis underline the versatility of CA platforms and indicate three types of platforms. The cross-cluster comparison shows that CA platforms range from high-end offerings (cluster 1), mainly by large technology providers such as IBM or Microsoft that offer a variety of analytical features and options for integration as well as provide CAs both for speech- and for text-based communication, over mid-range general-purpose CA platforms (cluster 2) like pandorabots or Chatfuel that primarily focus on single platforms for deployment and require implementing an API for integration to highly standardized CA platforms (cluster 3) that offer mainly domain-specific CA with a limited set of functionality, such as SurveyBot or MobileMonkey. These archetypes and the underlying taxonomy contribute to theory in different ways. The taxonomy we developed extends the existing, basic classification of CAs according to communication mode and context [5] through the empirical observation of CA design platforms by adding further dimensions. These dimensions describe CAs in greater detail as the existing classification, for example by taking into account the implementation approach, integration capabilities or the intelligence a CA possesses, which in turn provides possibilities and constraints for implementing CAs based on design principles formulated in previous studies [5]. Furthermore, we provide an overview of the state-of-the-art of platforms for conversational agent design through the taxonomy and classified platforms that can be used in future design-oriented research on CAs. For example, studies that investigate empathetic behavior of chatbots in customer service, such as the work by Hu et al. [51], could select a platform that offers built-in sentiment analysis for text-based CA to design their CA. Thus, in the context of design-oriented research, this study contributes to the growing knowledge base on CA [52].

In addition to the aforementioned contributions, our study provides two main insights for practitioners that intend to design CAs. First, the taxonomy can be used to select a vendor for a specific use case, for example by defining the desired characteristics along the 11 dimensions and then choosing a suitable platform. For example, a company that seeks to design a text-based CA with multi language support, on-premise hosting, and built-in analytics functionality could select a platform such as inbenta, Creative Virtual or IBM Watson Assistant. Or, a company that would like their CA to specifically conduct text-based customer surveys on a single platform, Facebook, can use SurveyMonkey for their implementation. The cross-cluster comparison shows that CA platforms range from high-end offerings (archetype 1), mainly by large technology providers such as IBM or Microsoft that offer a variety of analytical features and options for integration as well as provide CAs both for speech- and for text-based communication, over mid-range general-purpose CA platforms (archetype 2) like pandorabots or Chatfuel that primarily focus on single platforms for deployment and require implementing an API for integration to highly standardized CA platforms (archetype 3) that offer mainly domain-specific CAs with a limited set of functionality, such as SurveyBot or MobileMonkey.

Second, the platform database and identified archetypes underline the wide spectrum of CA platforms ranging from basic text-based CAs for single platforms to high-end, adaptive CAs that integrate in existing systems and can communicate with customers both via speech and text. Thus, managers can use the archetypes to strategically decide what type of CA platform they require. Furthermore, the analysis revealed that some platforms address different departments. Whereas multiple platforms can directly be used for design by the department that intends to introduce a CA, such as marketing and sales, as they deliver it based on simple modeling of typical conversation flows and convenient hosting in the cloud, other platforms address and require the IT department to customize, integrate and deploy their solutions.

5.2 Limitations and Opportunities for Future Research

Our study is not free of limitations and offers opportunities for future studies. First, the taxonomy that was developed both from existing CA literature and empirical data (i.e. the platforms in our database) cannot be considered comprehensive in terms of explaining platforms in detail but is helpful for understanding and delineating CA platforms as shown our analysis. As Nickerson et al. [41] highlight a taxonomy can never be perfect but is at best useful to explain the nature of objects under study. We initially demonstrated the usefulness of our taxonomy, but it can benefit from validation and expansion in future studies. A second limitation is that some dimensions might mutually exclude one another. We did not systematically identify these inter-dependencies in our work, yet it would be useful to address this point in the future. The third limitation results from the market dynamics that exist with regard to CA platforms. Present acquisitions, such as Motion.AI acquired by HubSpot, underline that the current CA platform landscape is subject to change which in turn limits the validity of our analysis over time. Similarly, CA platforms might add different functionality over time and provide new interfaces to enterprise software which would reduce the accuracy of our database. However, as the cluster analysis indicated a rather equal distribution of platforms to cluster, we would argue that the three typical CA platforms will still remain applicable even in the light of acquisitions and feature changes.

We suggest two main opportunities for future research: First, the taxonomy created in this paper can be evaluated in the field with organizations that plan to introduce CA for innovation or automation. Incorporating the views from organizations that seek to introduce a CA can be useful to validate and potentially extend the dimensions or characteristics in the taxonomy. Second, engaging with organizations introducing CA can also be helpful to reach a better understanding regarding the reasons for or against selecting specific archetypes as well as with regard to different characteristics. For example, comparing the two implementation approaches modeling of conversation flows with training of a CA based on existing and labeled data (supervised learning) concerning the impact on CA performance is a promising research endeavor not only in the context of CA, but also within the broader spectrum of innovative approaches for task or process automation.

6 Conclusion

In this study, we set out to develop a taxonomy of CA platforms (RQ1) and identify their archetypes (RQ2) in order to better understand the variety of platforms to design natural language agents for organizations. Based on existing CA literature as well as the analysis of 51 platforms, we derived a taxonomy with 11 dimensions which describes CA platform characteristics alongside their implementation and hosting approaches, pricing models, analytical features, and options for enterprise software integration. Afterwards, we empirically identified three archetypes of CA platforms with different ranges of functionality. Our work contributes an overview of the state-of-the-art of platforms for CA design and outlines possibilities and constraints for the implementation of design knowledge on conversational agents. In addition, our results can practically guide CA platform selection through the analysis of platforms based on the taxonomy and outlining aspects to be considered in the design process, such as the need for multi-language support or built-in sentiment analysis.

References

1. Brynjolfsson, E., McAfee, A.: The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. Norton & Company, New York, USA (2016).
2. Chakrabarti, C., Luger, G.F.: Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. Expert Syst. Appl. 42, 6878–6897 (2015).
3. Meuter, M.L., Ostrom, A.L., Roundtree, R.I., Bitner, M.J.: Self-Service Technologies:Understanding Customer Satisfaction with Technology-Based Service Encounters. J. Mark. 64, 50–64 (2000).
4. Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W.: Virtual customer service agents: Using social presence and personalization to shape online service encounters. J. Comput. Commun. 19, 529–545 (2014).
5. Gnewuch, U., Morana, S., Maedche, A.: Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–13. , Seoul, Korea (2017).
6. Jacobs, I., Powers, S., Seguin, B., Lynch, D.: The Top 10 Chatbots For Enterprise Customer Service. Forrester Res. (2017).
7. 80% of businesses want chatbots by 2020, https://www.businessinsider.de/80-of-businesses-want-chatbots-by-2020-2016-12?r=US&IR=T.
8. Wünderlich, N. V., Paluch, S.: A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–11. , Seoul, Korea (2017).
9. Morana, S., Friemel, C., Gnewuch, U., Maedche, A., Pfeiffer, J.: Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz. Wirtschaftsinformatik Manag. 5, 42–51 (2017).
10. Saffarizadeh, K., Boodraj, M., Alashoor, T.M.: Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 0–12. , Seoul, Korea (2017).
11. Schroeder, J., Schroeder, M.: Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). , Waikoloa , Hawaii, USA (2018).
12. Abul, M., Siddike, K., Spohrer, J., Demirkan, H., Kohda, Y.: People's Interactions with Cognitive Assistants for Enhanced Performances. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). pp. 1640–1648. , Waikoloa Village, Hawaii, USA (2018).
13. Oracle: Can Virtual Experiences Replace Reality? The future role for humans in delivering customer experience, (2016).
14. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics. Inf. Manag. 54, 545–559 (2017).
15. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Case study-Embodied virtual agents: An analysis on reasons for failure. J. Retail. Consum. Serv. 19, 605–612 (2012).
16. botnerds, http://botnerds.com/types-of-bots/.
17. Gnewuch, U., Morana, S., Adam, M.T.P., Maedche, A.: Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: Proceedings of the European Conference on Information Systems (ECIS). , Portsmouth, United Kingdom (2018).

18. Schuetzler, R.M., Grimes, G.M., Giboney, J.S.: An Investigation of Conversational Agent Relevance, Presence, and Engagement. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–10. , New Orleans, USA (2018).
19. McTear, M., Callejas, Z., Griol, D.: The Conversational Interface: Talking to Smart Devices. Springer Publishing Company, Basel, Switzerland (2016).
20. Berg, M.M.: Modelling of Natural Dialogues in the Context of Speech-based Information and Control Systems, http://macau.uni-kiel.de/receive/dissertation_diss_00016739, (2014).
21. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM. 9, 36–45 (1966).
22. Knijnenburg, B.P., Willemsen, M.C.: Inferring Capabilities of Intelligent Agents from Their External Traits. ACM Trans. Interact. Intell. Syst. 6, 1–25 (2016).
23. Berg, M.: NADIA: A Simplified Approach Towards the Development of Natural Dialogue Systems. In: Natural Language Processing and Information Systems. pp. 144–150. Springer International Publishing (2015).
24. Lee, C., Jung, S., Kim, S., Lee, G.G.: Example-based dialog modeling for practical multi-domain dialog system. Speech Commun. 51, 466–484 (2009).
25. Shawar, B.E., Atwell, E.: Chatbots: Are they really useful? In: LDV-Forum. pp. 29–49 (2007).
26. Hill, J., Randolph Ford, W., Farreras, I.G.: Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. Comput. Human Behav. 49, 245–250 (2015).
27. Sarikaya, R.: The technology behind personal digital assistants: An overview of the system architecture and key components. IEEE Signal Process. Mag. 34, 67–81 (2017).
28. Medium: 10 Real Examples How Brands are Using Chatbot for Customer Service, https://medium.com/the-mission/10-real-examples-how-brands-are-using-chatbot-for-customer-service-4fbb5e4617f3.
29. Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., Patton, M.W.: Embodied Conversational Agent-Based Kiosk for Automated Interviewing. J. Manag. Inf. Syst. 28, 17–48 (2011).
30. Cleverbot, https://www.cleverbot.com.
31. Mitsuku, http://www.mitsuku.com.
32. Vassos, S., Malliaraki, E., Dal Falco, F., Di Maggio, J.: Art-Bots: Toward Chat-Based Conversational Experiences in Museums. In: Proceedings of the 9th International Conference on Interactive Digital Storytelling. pp. 433–437. , Los Angeles, USA (2016).
33. Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A conversational agent as museum guide - Design and evaluation of a real-world application. In: Proceedings of Intelligent Virtual Agents: 5th International Working Conference. pp. 329–343. , Kos, Greece (2005).
34. Miner, A., Chow, A., Adler, S., Zaitsev, I., Tero, P., Darcy, A., Paepcke, A.: Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect. In: Proceedings of the International Conference on Human Agent Interaction, Singapore (2016).
35. Tiwana, A.: Platform Desertion by App Developers. J. Manag. Inf. Syst. 32, 40–77 (2015).
36. Chatfuel, https://chatfuel.com.
37. Twyla, https://www.twylahelps.com.
38. Glass, R.L., Vessey, I.: Contemporary Application-Domain Taxonomies. IEEE Softw. 12, 63–76 (1995).
39. McKnight, D.H., Chervany, N.L.: What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. Int. J. Electron. Commer. 6, 35–59 (2001).

40. Williams, K., Chatterjee, S., Rossi, M.: Design of emerging digital services: A taxonomy. Eur. J. Inf. Syst. 17, 505–517 (2008).
41. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. Eur. J. Inf. Syst. 22, 336–359 (2013).
42. Eickhoff, M., Muntermann, J., Weinrich, T.: What do FinTechs actually do ? A Taxonomy of FinTech Business Models. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–19. , Seoul, Korea (2017).
43. Varsheny, U., Nickerson, R.C., Muntermann, J.: Taxonomy Development in Health-IT. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–10. , Chicago, USA (2013).
44. Yang, A., Varshney, U.: A Taxonomy for Mobile Health Implementation and Evaluation. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–10. , Dublin, Ireland (2016).
45. Land, L., Smith, S., Pang, V.: Building a Taxonomy for Cybercrimes. In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS). pp. 34–38. , Jeju, Korea (2013).
46. Nickerson, R.C.: A Taxonomy of Collaborative Applications. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 560–562. , Indianapolis, USA (1997).
47. Remane, G., Hanelt, A., Nickerson, R.C., Tesch, J.F., Kolbe, L.M.: A Taxonomy of Carsharing Business Models. Proc. Int. Conf. Inf. Syst. 1–19 (2016).
48. Kaufman, L., Rousseeuw, P.J.: Introduction. In: Finding Groups in Data: An Introduction to Cluster Analysis. pp. 1–67 (1990).
49. Punj, G., Stewart, D.W.: Cluster Analysis in Marketing Research: Review and Suggestions for Application. J. Mark. Res. 20, 134–148 (1983).
50. Landau, S., Everitt, B.: A handbook of statistical analyses using SPSS. (2004).
51. Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., Akkiraju, R.: Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 1–12. , Montréal, Canada (2018).
52. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. Manag. Inf. Syst. Q. 28, 75–105 (2004).

Measuring Service Encounter Satisfaction with Customer Service Chatbots using Sentiment Analysis

Jasper Feine[1], Stefan Morana[1], and Ulrich Gnewuch[1]

[1] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing (IISM), Karlsruhe, Germany
{jasper.feine,stefan.morana,ulrich.gnewuch}@kit.edu

Abstract. Chatbots are software-based systems designed to interact with humans using text-based natural language and have attracted considerable interest in online service encounters. In this context, service providers face the challenge of measuring chatbot service encounter satisfaction (CSES), as most approaches are limited to post-interaction surveys that are rarely answered and often biased. As a result, service providers cannot react quickly to service failures and dissatisfied customers. To address this challenge, we investigate the application of automated sentiment analysis methods as a proxy to measure CSES. Therefore, we first compare different sentiment analysis methods. Second, we investigate the relationship between objectively computed sentiment scores of dialogs and subjectively measured CSES values. Third, we evaluate whether this relationship also exists for utterance sequences throughout the dialog. The paper contributes by proposing and applying an automatic and objective approach to use sentiment scores as a proxy to measure CSES.

Keywords: online customer service, chatbot, sentiment analysis, service encounter satisfaction, correlation analysis

1 Introduction

Digital communication technologies have become an integral part for organizations to interact with their customers [1]. Many companies offer online services via live chat interfaces, which enable customers to directly interact with customer service employees [2]. This type of text-based service encounter is a cost effective service solution and often the preferred way of communication for young people [3]. One technology which is often deployed to assist service employees in online service encounters are chatbots [1]. Chatbots are software-based systems designed to interact with humans via text-based natural language [4, 5] and can be found across industries (e.g., airlines, energy provider). Gartner predicts that by 2020, 25% of all customer services organizations will integrate this technology [6].

Despite their great potential, many customer service chatbots did not meet customer expectations and led to service failures [7]. As a result, many service providers retired their chatbots, as unsatisfactory online service encounters have negative effects on

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

word-of-mouth, loyalty, and intention to repurchase a product [8]. Ignoring customer frustrations can strongly impede the performance of customer service encounters and carries the risks that the service chatbot is perceived as cold, socially indebt, untrustworthy, and incompetent [9]. Therefore, service providers should identify service encounters that were below customer's expectations [10] and trigger service recovery procedures (e.g., offering compensation). Such procedures can help to recover from almost any service failure and increase trust, perception of fairness, and service experience [10]. However, most approaches to identify dissatisfied customers in a text-based online environment (e.g., chat, social media) are limited to post-interaction surveys [11]. This is problematic as self-reported data can hardly be retrieved during an interaction, is influenced by various biases [12], and only few users are willing to provide this kind of information [8, 11]. Therefore, we propose that an automated method to measure chatbot service encounter satisfaction (CSES) during a customer-chatbot interaction could help service providers to deal with these issues.

To develop such a method, we want to take advantage of the fact that written text is associated with a person's thoughts, emotions and motivations [13–15]. Humans write differently when they are happy or frustrated and thus, written text by itself conveys much information about a human [14]. Users who are less happy with a chatbot use less assent, fewer positive, and more anger-related words and thus, express more negative sentiments [16]. The analysis of such opinionated text can provide valuable information about the user as opinions are "*key influencers of our behaviors*" [17, p. 2] and "*sentiment and tonal polarity are inherent properties of human-human communication and interaction*" [18, p. 1367].

As a manual analysis of expressed polarity in written text does not scale well to larger datasets [19], automated sentiment analysis methods have been developed. These methods are capable of automatically extracting positive or negative polarity expressed in written text [20]. Moreover, current sentiment analysis methods have been found to be very accurate and thus, seem to be a valid approach [21, 22]. However, research has rarely applied sentiment analysis in human-computer interaction (HCI) so far [20, 23]. Most HCI studies focus on auditory and visual signals of humans as these transmit the majority of communication-related information [20]. Moreover, most sentiment analysis studies focus on the method itself [23]. As a result, there is a lack of understanding on how to apply sentiment analysis in online chatbot service encounters to obtain valuable information about the user and her/his CSES. Therefore, we investigate the application of sentiment analysis methods for chatbots in online service encounters by drawing on research that text-based communication by itself is rich in informative signals [24] and that written language is influenced by emotions, intentions, and thoughts [13–15]. More specifically, we argue that sentiment analysis of dialog data can be used as an easy-to-use and objective proxy to measure CSES. Therefore, our research project addresses the following research question:

How to measure service encounter satisfaction with a chatbot using sentiment analysis methods?

To address this research question, we first compare different sentiment analysis methods on an empirical level by analyzing the calculated sentiment scores for two datasets. Next, we test for a potential correlation between sentiment scores and CSES values that were measured using a survey-based approach in an online experiment. In doing so, we first investigate this potential relationship on a dialog level and second on an utterance level (i.e., single messages). This paper contributes by proposing and applying an automatic and objective approach to use sentiment scores as a proxy to measure CSES. Our proposed approach enables researchers and practitioners, such as online customer service providers, to objectively and automatically retrieve valuable information after and during an online service encounter.

2 Related Work

2.1 Customer Service Chatbots

Recent advances in technology and great business potential have led to an increased interest in the development of conversational agents [5, 25]. Conversational agents are software-based systems designed to converse with a user via natural language [4, 5]. Thereby, the user interacts with the conversational agent in a natural dialog and does not use a predefined set of keywords or command phrases [4]. They can offer both speech- and text-based interfaces and can also be visualized and animated (i.e., embodied conversational agent) [4]. Conversational agents that interact with the user primarily via a text-based interface are often referred to as chatbots [5]. Chatbots can be deployed on various communication channels, such as instant messaging platforms (e.g., Line, Telegram, WeChat), websites, or on social media (e.g., Facebook, Twitter) and are accessible from various devices (e.g., PCs, mobile phones) [4]. Since Weizenbaum developed the first chatbot named ELIZA in 1966, much research has been conducted and various chatbots have been deployed across industries [4, 5].

One of the reasons why both research and practice are increasingly using this technology is the fact that chatbots interact in a human-like interaction style (i.e., use natural language) and offer great business potential (i.e., 24/7 availability at lost costs) [4]. Therefore, chatbots are increasingly implemented in online service encounters as many companies communicate with their customers via live-chats on their website or on social media platforms [1, 2]. Chatbots could help to automate online customer service, save costs, and enhance online experience [1, 26]. For example, instead of a customer calling or chatting with a service employee, customers are now communicating with a service chatbot [26]. In addition, chatbots can also take the role of first tier support agents and assist customer service employees. Therefore, chatbots can first start an online service encounter and then seemingly handover the conversation to a human agent when required. This can lead to a great reduction of routine requests usually handled by service employees.

2.2 Chabot Service Encounter Satisfaction

Satisfaction is an often applied construct in information systems (IS) research to evaluate the success and effectiveness of a system and it is particularly critical for the success of service systems [27]. It reflects whether customers perceive a service as pleasurable with regard to its consumption-related fulfilment [8]. High customer satisfaction values are important to achieve long-term success, especially in highly competitive markets, and therefore should have priority for any organization [8, 28].

Customer satisfaction is strongly impacted by the service encounter satisfaction, which refers to the post-consumption evaluation of a service encounter [29, 30]. A successful service encounter makes a company's product incrementally more effective and easier to use [28], influences the customer's choice independent whether a service is provided offline or online [31], and is linked to several desired outcomes such as word-of-mouth, loyalty, and intention to repurchase a product [8, 32]. Thus, service encounter satisfaction is a critical indicator for any organization [8, 28].

Service encounter satisfaction is influenced by several antecedents such as the customization and flexibility in service encounter, effective service recovery when failures occur, and spontaneous delights (i.e., pleasing experiences customers do not expect) [32]. In addition, various design elements of a chatbot influence the CSES such as verbal communication cues (i.e., being polite, responsive, and show mutual understanding), level of expertise (i.e., a core attribute of a service employee), or visual cues (i.e., such as an avatar) [29]. In an online context, the measurement of CSES is often limited to follow-up surveys [11, 29]. Thus, CSES cannot be retrieved in real-time, is often biased, and the surveys are only answered by a few users [11, 12]. Moreover, customers have a general "*reluctance to share their sentiments with firms*" [8, p. 359] and thus, companies are often not able to react fast enough to dissatisfied customers using service recovery procedures [10]. Failing to recover can result in lost customers, negative word of mouth, decreased loyalty, and less profits [28, 32].

2.3 Sentiment Analysis Methods

A common method within the natural language understanding literature is the analysis of opinions and sentiments expressed in written text. This becomes meaningful as research has shown that written text is clearly impacted by the user's emotions, intentions, and thoughts [13–15]. Consequently, written text says something about us and can be used as a proxy for information about the author. Therefore, various methods have been developed to analyze the opinions and sentiments expressed in written text [21]. These methods are named and defined in many different ways (e.g., sentiment analysis, opinion mining, see [33]). As it is the most common name, we follow [17, 33] and define sentiment analysis as the computational analysis of written language to identify the user's perceived positive or negative valence towards a certain entity (e.g., product, service, event). Sentiment analysis has recently witnessed great attention, because of the large availability of opinion-rich resources on the Internet (e.g., online reviews) and advances in artificial intelligence [17]. Consequently, many of the major technology companies offer sentiment analysis solutions (e.g., IBM, Google) and also various open source solutions are available (see [21]). This led to the development of many available and precise methods (see [20, 21]).

Sentiment analysis methods can be generally distinguished into two broad but also overlapping approaches, namely the application of semantic rules or statistical methods [20]. Methods of the first category compare sentiment-related expressions with sentiment lexicons that contain the semantic orientation of words [34]. One of the greatest challenges of these methods is that the semantic orientation of individual words does not necessarily correspond to the contextual polarity of the whole sentence [34]. Therefore, it is necessary to extract additional linguistic patterns of the text by conducting morpho-syntactic text analyses (i.e., wordform, lemma, part of speech tags) [20]. Too specific extraction patterns, however, limit the application range to a specific domain. Methods of the second, more recently applied category use unsupervised or supervised machine learning algorithms including support vector machines and Bayes classifiers [20]. These methods enable the development of more generic models, but require labeled data for training purposes. Consequently, the quality of such models is heavily influenced by the reliability of sentiment annotations [20].

Today's applications of sentiment analysis are manifold. Sentiment analysis can be used to predict the success of political campaigns [35], identify interaction problems within a conversation corpus [11], or even to scan the dark web in an intelligence context [36]. Nevertheless, only a few studies analyzed sentiments in a chatbot context yet as most studies are focusing on the method itself [20, 23]. One reason is the difficulty to classify rather short informal chat messages, which include a high degree of language creativity, spelling mistakes, and the expression of sentiments without real intentions [19]. Another reason are the differences and ambiguities in human mood coding which make it difficult to create a gold standard [37] and thus, it is difficult to develop user-independent prediction models [38]. However, some related research has already applied sentiment analysis to infer the customer satisfaction from product reviews for shopping websites and mobile services [23, 39].

3 Research Method

To answer our research question, we first selected suitable dialog corpora and sentiment analysis methods to run our analyses. Then, we defined a three-step research approach to analyze the corpora in order to answer our research question.

3.1 Dialog Corpora and Sentiment Methods

First, we collected one dialog corpus from an online experiment in a customer service context [40]. The participants (n = 79, mean age = 28.835, SD age = 6.388) were given a fictive mobile phone bill and the experimental task was to find a more suitable mobile phone plan through interacting with a customer service chatbot. The chatbot asked several consumption-related questions and was capable of responding interactively to given user queries. After the interaction, all participants were asked to complete a questionnaire measuring CSES using an established measurement instrument on a 7-point Likert scale [29]. The construct displayed a sufficient composite reliability (CR) above 0.8 (CR = 0.814) and the average variance extracted was above 0.5. All measurement items had factor loadings above 0.7 and the mean CSES value was 4.924 (SD = 1.179). The complete experiment, all dialogs, as well as the questionnaire were in English. The complete dialog corpus consists of 79 user dialogs with a total of 1416 user utterances. We removed 353 utterances because they consisted of only mobile contract related numbers. The final corpus included 79 dialogs and a total of 1063 utterances with an average of 13.456 utterances per dialog (SD = 8.312). We refer to this dialog corpus as "ExpCorpus" in the remainder of this paper.

In addition to ExpCorpus, we used a second, publicly available dialog corpus (without CSES values) in order to have a greater basis for the comparison of different sentiment analysis methods. Therefore, we selected the "ConvAI" dialog corpus [41]. 500 volunteers chatted with ten chatbots and the dialog set is freely available as a JSON-File. The dataset includes 2778 dialogs from which we excluded 441 human-human dialogs, 102 empty dialogs, 54 bot only dialogs, and one numbers-only dialog. This resulted in the extraction of 2180 human-chatbot dialogs, which were neither empty nor contained only numbers. Finally, we extracted all 12482 human written utterances. We refer to this dialog corpus as "ConvAI" in the remainder of this paper.

To select appropriate sentiment analysis methods for our study, we reviewed two benchmark analyses [21, 42]. We followed the benchmark analysis of [21], which compared 24 open source methods, as well as the benchmark analysis of [42], which also included sentiment analysis methods from major technology companies (e.g., IBM, Microsoft). The benchmark analyses reveal that there is no superior sentiment analysis method because all tools perform differently depending on the specific context they are applied on or depending on the corresponding data source on which they were trained [21]. Consequently, both benchmarks reveal several suitable methods depending on the respective context and the training data [21]. The benchmark of [21] reveals that two of the best sentiment analysis methods providing numerical polarity for negative, neutral, and positive sentiments are VADER [43] and AFINN (i.e., an extension of ANEW [44]) [21]. VADER and AFINN are rule-based sentiment analysis methods, which use rules and heuristics to match the analyzed texts to sentiment lexicons. Both lexicons were developed and trained on social media content and Twitter data [21, 43]. The benchmark analysis of [42] reveals that the sentiment analysis methods by IBM Watson, Google Cloud, and Microsoft Azure perform best with varying types of datasets [42]. These sentiment analysis methods leverage machine learning classification algorithms in order to predict the sentiment score. Therefore, all three providers trained their algorithms on an extensive body of sentiment annotated text databases [42]. To cover both types of sentiment analysis techniques, namely semantic rules and statistical methods [20], we selected the following methods for our study: two open source methods using rule-based sentiment analysis methods (i.e., VADER, AFINN) and three commercial methods using machine learning classification algorithms (i.e., IBM Watson, Google Cloud, and Microsoft Azure). We calculated the sentiment scores for each of the open source methods using the web service ifeel 2.0 provided by [22] and for each of the commercial methods using their Node.js APIs.

3.2 Research Approach

In this section, we present our three-step research approach (see Table 1) to answer our research question and to investigate the potential correlation between sentiments and CSES. All analyses were conducted using R 3.5.0.

Table 1. Research approach

Step	Research method	Dialog corpora	Sentiment methods
1.	Comparison of sentiment analysis methods	ConvAI (dialog & utterance level), ExpCorp (dialog & utterance level)	VADER, AFINN, IBM, Microsoft, Google
2.	Correlation analysis between sentiment scores and CSES values	ExpCorp (dialog level)	VADER, AFINN, IBM, Microsoft, Google
3.	Exploratory analysis of sentiment scores and CSES values	ExpCorp (utterance level)	IBM

In the first step, we compared all selected sentiment methods because the accuracy of sentiment analysis method are highly context and data dependent. Therefore, we investigated whether sentiment scores from each tool are similar on a dialog and utterance level by calculating the sentiment scores for each dialog and each single utterance of both corpora with all five methods. Next, we tested for potential correlations among the five sentiment scores. We do this analysis on a sentence and utterance level as sentiment analysis methods seem to perform better on "*carefully authored, lengthier content, but often struggle when faced with informal online communication*" [19, p. 318]. Consequently, we assume that some sentiment methods may struggle to predict the sentiment score of rather short utterance level and that the methods perform quite differently on both levels.

In the second step, we tested for a correlation between sentiment scores and CSES values. Therefore, we standardized the sentiment scores to -1 (i.e., negative) and +1 (i.e., positive) and subsequently tested for a correlation between sentiment scores (from all five methods) and CSES values using the dialogs and satisfaction data of ExpCorpus. By doing this, we aimed to reveal whether sentiment scores are a valid proxy for CSES values.

In the third step, we investigated the minimum number of utterances required to show a correlation between sentiment scores and CSES values. For this analysis, we used IBM's sentiment method because it yielded the highest correlation in the previous step. Therefore, we extracted utterance sequences of each dialog, calculated their sentiment scores, and tested for a correlation between sentiment scores and CSES values. Next, we investigated whether these findings also hold for utterance sequences throughout the whole dialog. This analysis provides insights whether sentiment scores can be used as a proxy for CSES during a customer service encounter.

4 Results

Step 1: Comparison of Sentiment Analysis Methods

In the first step, we started our analysis by comparing the calculated sentiment scores of selected sentiment analysis methods for both dialog corpora (ConvAI and ExpCorpus). Table 2 contains the correlation analysis between sentiment scores of both corpora for each dialog and single utterance calculated by all five methods.

Table 2. Pearson correlation analyses among sentiment scores of different methods

Corpus	Method	AFINN	VADER	IBM	Microsoft	AFINN	VADER	IBM	Microsoft
ConvAI	AFINN	-				-			
	VADER	.605***	-			.322***	-		
	IBM	.385***	.317***	-		.387***	.357***	-	
	Microsoft	.368***	.356***	.533***	-	.300***	.311***	.604***	-
	Google	.369***	.295***	.504***	.395***	.414***	.388***	.600***	.497***
		n = 2180 dialogs				n = 12482 utterances			
ExpCorpus	AFINN	-				-			
	VADER	.719***	-			.597***	-		
	IBM	.508***	.512***	-		.505***	.169***	-	
	Microsoft	.473***	.467***	.600***	-	.383***	.201***	.615***	-
	Google	.516***	.366***	.521***	.537***	.653***	.439***	.625 ***	.487***
		n = 79 dialogs				n = 1063 utterances			
*** $p < .001$									

The results reveal that sentiment scores of dialog data from both corpora are at least moderately positively correlated with each other [45] (ConvAi $.295 \leq r \leq .605$, $n = 2180$, $p < .001$, ExpCorpus $.366 \leq r \leq .719$, $n = 79$, $p < .001$). The strongest correlation for ConvAi dialogs were identified between VADER's and AFINN's sentiment scores ($r = .605$, $n = 2180$, $p < .001$) and the weakest between Vader's and Google's sentiment scores ($r = .295$, $n = 2180$, $p < .001$). The strongest correlation for ExpCorpus was again identified between VADER's and AFINN's sentiment scores ($r = .719$, $n = 79$, $p < .001$) and the weakest one between Vader's and Google's sentiment scores ($r = .366$, $n = 2180$, $p < .001$). All sentiment scores on an utterance level were significantly positively correlated, but some correlations were weaker among some methods than they were on a dialog level ($.169 \leq r \leq .653$, $p < .001$). All in all, the findings reveal that sentiment methods using similar methodologies to identify the expressed polarity in a given text provide rather similar results. Thus, methods using semantic rules such as VADER and AFINN are strongly correlated on a dialog level. Moreover, methods using machine classification algorithms such as IBM's, Microsoft's, and Google's methods are at least moderately correlated on a dialog and utterance level.

Step 2. Correlation Analysis Between Sentiment Scores and CSES Values

In the second step, we tested for a correlation between sentiment scores and CSES values using the dialogs and CSES values of ExpCorpus. The results and the corresponding scatterplots are displayed in Figure 1. The analysis reveals a significant moderate to strong correlation between sentiment scores (from all five methods) and CSES values ($.405 \leq r \leq .513$, $n = 79$, $p < .001$) [45]. Thus, we conclude that there is a moderate positive correlation between sentiment scores and CSES values for four sentiment analysis methods and a strong positive correlation for IBM's sentiment method ($r = .513$, $n = 79$, $p < .001$) [45]. Moreover, it becomes visible that sentiment scores seem to be primarily a better predictor for positive than for negative CSES values. Moreover, semantic rule based algorithms seem to calculate sentiment scores of the dialogs generally more positive.

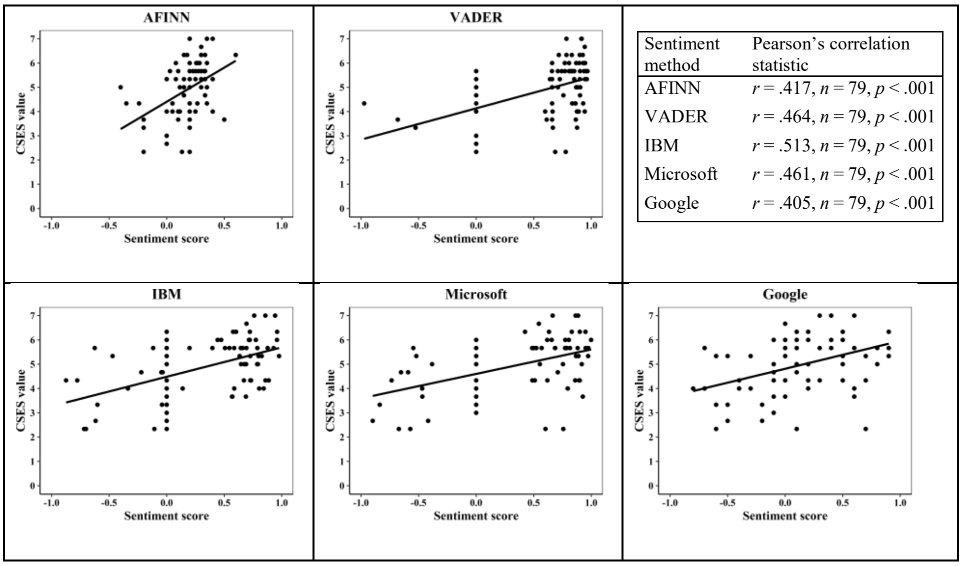


Figure 1. Correlation analyses between sentiment scores (of dialogs) and CSES values for ExpCorpus

Step 3: Exploratory Analysis of Sentiment Scores and CSES Values

In the third step, we investigated the minimum number of utterances required to show a significant positive correlation between sentiment scores and CSES values. Therefore, we combined the first ten utterances (u_i, i = 1, … ,10) into ten different utterance sequences (US_i, i = 1, …, 10), calculated their sentiment scores, and tested for a correlation with CSES values. The results are summarized in Table 3.

Table 3. Correlation analyses between sentiment scores (of utterances sequences) and CSES values for ExpCorpus

Analysed utterance squence	Included dialogs	Included utterances	Included words	Pearson' s correlation statistic
$US_1 = \{u_1\}$	79	79	739	$n = 79, r = .018, p = .872$
$US_2 = \{u_1, u_2\}$	79	158	1086	$n = 79, r = .133, p = .244$
$US_3 = \{u_1, u_2, u_3\}$	79	237	1360	$n = 79, r = .234, p = .038$
$US_4 = \{u_1, …, u_4\}$	79	316	1574	$n = 79, r = .251, p = .026$
$US_5 = \{u_1, …, u_5\}$	79	395	1779	$n = 79, r = .372, p < .001$
$US_6 = \{u_1, …, u_6\}$	75	450	1989	$n = 75, r = .437, p < .001$
$US_7 = \{u_1, …, u_7\}$	74	518	2159	$n = 74, r = .480, p < .001$
$US_8 = \{u_1, …, u_8\}$	68	544	2350	$n = 68, r = .443, p < .001$
$US_9 = \{u_1, …, u_9\}$	58	522	2495	$n = 58, r = .506, p < .001$
$US_{10} = \{u_1, …, u_{10}\}$	46	460	2633	$n = 46, r = .503, p < .001$
All dialogs with all utterances	79	1060	3431	$n = 79, r = .513, p < .001$

Please note that not all dialogs included up to ten user utterances. As a consequence, the number of analyzed dialogs decreases with increasing sequence length. The last row analyzes all dialogs including all utterances of each dialog.

The analysis reveals that the sentiment scores of US_1 and US_2 have no significant correlation with the CSES values. However, the correlation increases with an increasing number of utterances combined in each sequence. Our results show a significant weak positive correlation between sentiment scores and CSES values after the analysis of the first three utterances ($r = .234$, $n = 79$, $p = .038$). Moreover, we revealed a significant moderate positive correlation ($r = .372$, $n = 79$, $p < .001$) after the analysis of the first five utterances [45]. To provide a greater understanding of these findings, Table 4 provides some exemplary utterance sequences, their sentiment scores, and the measured CSES values.

Table 4. Exemplary utterance sequences including the first three utterances

Utterance sequence	Sentiment score	CSES value
{"*Hi*", "*Nice to meet you. I'm interested in a cheaper phone plan. Can you help me?*", "*I think it is SuperMobile*"}	0.769	6
{"*Hey, I'm currently on the mobile phone plan Yellow Basic 1000 and I received an unexpectedly high mobile phone billl last month.*", "*Are there any better mobile phone plans for me?*", *"It's SuperMobile Yellow Basic 1000"*}	0.488	6
{"*My bill is too high*", "*Help me to find a new mobile phone plan*", "*I dont know*"}	-0.566	4,333

Having shown that at least the first three utterances of a dialog are required to find a significant positive correlation between sentiment scores and CSES values, we further investigated whether this correlation can also be found for all utterance sequences throughout the whole dialogs. Therefore, we extracted all consecutive utterance sequences consisting of three or five utterances within the first ten utterances of each dialog. This extraction resulted in eight consecutive utterance sequences for dialogs that were at least that long (e.g., *Seq-1* = {*u1*, *u2*, *u3*}, *Seq-2* = {*u2*, *u3, u4*}, *Seq-9* = {*u1*, *u2*, *u3*, *u4*, *u5*}). Then we calculated the sentiment scores and tested for a correlation between sentiment scores and CSES values (see Table 5).

Table 5. Correlation analyses between sentiment scores (of consecutive utterance sequences) and CSES values for ExpCorpus

Sequence	Included utterances	Pearson' s correlation statistic
Seq-1	{*u1, u2, u3*}	$n = 79$, $r = .234$, $p = .038$
Seq-2	{*u2, u3, u4*}	$n = 79$, $r = .243$, $p = .031$
Seq-3	{*u3, u4, u5*}	$n = 79$, $r = .289$, $p = .010$
Seq-4	{*u4, u5, u6*}	$n = 75$, $r = .350$, $p = .002$
Seq-5	{*u5, u6, u7*}	$n = 74$, $r = .410$, $p < .001$
Seq-6	{*u6, u7, u8*}	$n = 68$, $r = .267$, $p = .029$
Seq-7	{*u7, u8, u9*}	$n = 58$, $r = .501$, $p < .001$
Seq-8	{*u8, u9, u10*}	$n = 46$, $r = .501$, $p < .001$
Seq-9	{*u1, u2, u3, u4, u5*}	$n = 79$, $r = .372$, $p < .001$
Seq-10	{*u2, u3, u4, u5, u6*}	$n = 75$, $r = .407$, $p < .001$
Seq-11	{*u3, u4, u5, u6, u7*}	$n = 74$, $r = .436$, $p < .001$
Seq-12	{*u4, u5, u6, u7, u8*}	$n = 68$, $r = .377$, $p = .002$
Seq-13	{*u5, u6, u7, u8, u9*}	$n = 58$, $r = .503$, $p < .001$
Seq-14	{*u6, u7, u8, u9, u10*}	$n = 46$, $r = .325$, $p = .028$
Please note that not all dialogs included up to ten user utterances. As a consequence, the number of analyzed dialogs decreases with increasing utterance position.		

The analysis shows that sentiment scores of all utterance sequences throughout the whole dialog are positively correlated with the CSES values. All correlations are significant at least at a $p < .05$ level. The correlation strength varies among the different sequences between weak and strong correlation. However, the minimum and maximum value of the correlation strength is higher for sequences consisting of five consecutive utterances, which always had at least a moderate positive correlation with CSES values.

5 Discussion

In this paper, we investigate whether sentiment scores from textual input can be used as a proxy to measure CSES in a customer-chatbot interaction. Therefore, we followed a three-step research approach: first, we compared five sentiment analysis methods by testing the relation of sentiment scores from two dialog corpora. Second, we tested for a correlation between sentiment scores and CSES values. Third, we analyzed this correlation in detail at the utterance level. Results of step 1 reveal a significant positively correlation among sentiment scores from all selected sentiment analysis methods. Results of step two reveal that sentiment scores of complete dialogs are significantly positive correlated with the subjectively measured CSES values. Results of step three reveal that this relation is not only valid for the analysis of an entire dialog, but also for any sequences of at least three consecutive utterances throughout the entire dialog. Thus, we conclude that sentiment scores can be used as an automatic and objective proxy to measure CSES in an online service encounter. Therefore, our findings further contribute to existing research that states "*sentiment analysis corresponds surprisingly well with emotional self-report*" [15, p. 87].

The results of our analysis have implications for the design of customer service chatbots. As customers may express their frustrations in written language, future chatbots could continuously perform sentiment analyses and use sentiment scores as a proxy to identify dissatisfied customers (by analyzing at least three consecutive utterances). In this way, service providers can intervene to reduce the risk of service failures. For example, a customer service chatbot could recognize that the current conversation with a customer is turning towards a negative sentiment score. In this case, several strategies could be triggered. The chatbot could seamlessly handover the conversation to a trained human service agent, automatically trigger service recovery procedures, or express certain verbal utterances such as excuses [46, 47]. Research has shown that these immediate reactions can reduce the level of frustration [46] and can lead to an increased interaction length [47]. Furthermore, service providers can use this data in post-interaction analyses to retrieve valuable information about CSES. This information cannot only be used for service recovery, but also for identifying general weaknesses in the service quality of the chatbot.

Although we aimed to ensure a high rigor in our research, some limitations should be considered. First, many sentiment analysis methods exist and they all may evaluate a given text differently depending on the context and type of a message [21]. This becomes even more meaningful when applied to rather short and informal chat data. Therefore, a different selection of sentiment methods may have led to different results.

Consequently, we tried to minimize this risk by starting with a selection of five sentiment methods based on benchmarks and compared them with each other by applying them on two dialog corpora. Even though all sentiment analysis methods had a moderate to strong correlation to CSES, some sentiments methods were rather weak predictors for users having low CSES values. Therefore, it "*is important that researchers and companies perform experiments with different methods before applying a method*" [21, p. 27]. Second, we analyzed a dialog corpus, which measured the CSES using a post-interaction survey. However, data of a survey-based approach might be influenced by various biases [12]. To reduce this risk, we reviewed all dialogs and verified that participants followed the experimental task and did not answer with straight line responses. Third, we only analyzed the relationship between sentiments and CSES based on a dialog corpus from a hypothetical online service task (i.e., finding new plan) in a specific context (i.e., mobile contract) in one language (i.e., English). Therefore, it is unclear whether our findings also hold for other customer service tasks (e.g., book ticket) in other contexts (e.g., airlines) in other languages (e.g., German). Fourth, we conducted correlation analyses between sentiment scores and CSES values to reveal a correlation between these two variables. Even though we found a strong positive correlation and propose sentiment scores as a proxy for CSES values, this analysis does not provide the explanation for this relation and does not indicate a cause-and-effect relationship [48]. Thus, results need to be applied with care as we cannot predict CSES based on sentiments scores or vice versa [48].

Considering these limitations, we identify several avenues for future research. First, future work can replicate our analyses on additional dialog corpora from different contexts, doing different tasks, and in different languages. This could further strengthen the applicability of sentiment analysis as a proxy to measure CSES in several domains and languages. Second, future studies could investigate adaptive reaction strategies based on real-time analyses of at least three consecutive user utterances. This could enable chatbots to recognize user frustrations and supports the development of chatbots that act more socially [46, 47]. Moreover, future research could investigate the application of more trivial text analysis methods, such as word count and length of dialogs, as well as more complex methods, such as topic modelling, as proxies to predict customer satisfaction. Integrating these techniques into a chatbot can lead to even greater understanding of the user and enables more precise reactions by the chatbot.

6 Conclusion

In this paper, we investigate the application of sentiment analysis methods in an online service encounter with a chatbot and show that sentiment scores can serve as a proxy to measure CSES. This enables researchers and practitioners, such as online service providers, to objectively and automatically retrieve user information during and after an online service encounter. This information can be used not only to trigger service recovery procedures, but also to identify weaknesses in the service quality and to analyze the user in real-time. Therefore, our results contribute towards the design of user adaptive service chatbots.

References

1. Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wünderlich, N.V., Keyser, A. de: “Service Encounter 2.0”: An investigation into the roles of technology, employees and customers. Journal of Business Research 79, 238–246 (2017)
2. McLean, G., Osei-Frimpong, K.: Examining satisfaction with the experience during a live chat service encounter-implications for website providers. Computers in Human Behavior 76, 494–508 (2017)
3. Kowatsch, T., Nißen, M., Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., Wangenheim, F. von: The Impact of Interpersonal Closeness Cues in Text-based Healthcare Chatbots on Attachment Bond and the Desire to Continue Interacting: An Experimental Design. In: Twenty-Sixth European Conference on Information Systems (ECIS). Portsmouth, UK (2018)
4. McTear, M., Callejas, Z., Griol, D.: The Conversational Interface. Talking to Smart Devices. Springer International Publishing, Switzerland (2016)
5. Dale, R.: The return of the chatbots. Natural Language Engineering 22, 811–817 (2016)
6. Gartner: Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020, https://www.gartner.com/newsroom/id/3858564
7. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Case study—Embodied virtual agents. An analysis on reasons for failure. Journal of Retailing and Consumer Services 19, 605–612 (2012)
8. Oliver, R.L.: Satisfaction. A behavioral perspective on the consumer. McGraw Hill, New York (1997)
9. Brave, S., Nass, C.: Emotion in human-computer interaction. In: Jacko, J.A., Sears, A. (eds.) The human-computer interaction handbook, pp. 81–96. L. Erlbaum Associates Inc, NJ, USA (2002)
10. Holloway, B.B., Beatty, S.E.: Service Failure in Online Retailing: A Recovery Opportunity. Journal of Service Research 6, 92–105 (2003)
11. Xiang, Y., Zhang, Y., Zhou, X., Wang, X., Qin, Y.: Problematic situation analysis and automatic recognition for chinese online conversational system. In: Joint Conference on Chinese Language Processing. Wuhan, China (2014)
12. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: a critical review of the literature and recommended remedies. The Journal of applied psychology 88, 879–903 (2003)
13. Tausczik, Y.R., Pennebaker, J.W.: The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology 29, 24–54 (2010)
14. Nerbonne, J.: The Secret Life of Pronouns. What Our Words Say About Us. Literary and Linguistic Computing 29, 139–142 (2014)
15. Küster, D., Kappas, A.: Measuring Emotions Online: Expression and Physiology. In: Holyst, J.A. (ed.) Cyberemotions: Collective Emotions in Cyberspace, pp. 71–93. Springer International Publishing, Cham (2017)
16. Skowron, M., Rank, S., Theunis, M., Sienkiewicz, J.: The Good, the Bad and the Neutral: Affective Profile in Dialog System-User Communication. In: D’Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) Affective Computing and Intelligent Interaction, pp. 337–346. Springer, Berlin, Heidelberg (2011)
17. Liu, B.: Sentiment analysis and opinion mining. Synthesis lectures on human language technologies 5, 1–167 (2012)

18. Banchs, R.E.: On the construction of more human-like chatbots: Affect and emotion analysis of movie dialogue data. In: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Kuala Lumpur (2017)
19. Brooks, M., Kuksenok, K., Torkildson, M.K., Perry, D., Robinson, J.J., Scott, T.J., Anicello, O., Zukowski, A., Harris, P., Aragon, C.R.: Statistical Affect Detection in Collaborative Chat. In: Conference on Computer Supported Cooperative Work, pp. 317–328. ACM, New York, NY, USA (2013)
20. Clavel, C., Callejas Z.: Sentiment Analysis: From Opinion Mining to Human-Agent Interaction. IEEE Transactions on affective computing 7, 74–93 (2016)
21. Ribeiro, F.N., Araújo, M., Gonçalves, P., André Gonçalves, M., Benevenuto, F.: SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Science 5, 23 (2016)
22. Diniz, J.P., Bastos, L., Soares, E., Ferreira, M., Ribeiro, F., Benevenuto, F.: ifeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis. In: 10th international AAAI conference on weblogs and social media. Cologne, Germany (2016)
23. Kang, D., Park, Y.: Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications 41, 1041–1050 (2014)
24. Walther, J.B., Parks, M.R.: Cues filtered out, cues filtered in. In: Knapp, M.L., Daly, J.A. (eds.) Handbook of Interpersonal Communication, pp. 529–563. SAGE, Thousand Oaks, CA, USA (2002)
25. Maedche, A., Morana, S., Schacht, S., Werth, D., Krumeich, J.: Advanced User Assistance Systems. Business & Information Systems Engineering 58, 367–370 (2016)
26. Gnewuch, U., Morana, S., Maedche, A.: Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: Proceedings of the 38th International Conference on Information Systems (ICIS). AISel, Seoul (2017)
27. Au, N., Ngai, E.W.T., Cheng, T.E.: A critical review of end-user information system satisfaction research and a new research framework. Omega-International Journal of Management Science 30, 451–478 (2002)
28. Jones, T.O., Sasser, W.E.: Why satisfied customers defect. Harvard Business Review 73, 88-& (1995)
29. Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W.: Virtual Customer Service Agents. Using Social Presence and Personalization to Shape Online Service Encounters. Journal of Computer-Mediated Communication 19, 529–545 (2014)
30. Caruana, A.: Service loyalty. European Journal of Marketing 36, 811–828 (2002)
31. Shankar, V., Smith, A.K., Rangaswamy, A.: Customer satisfaction and loyalty in online and offline environments. International Journal of Research in Marketing 20, 153–175 (2003)
32. Bitner, M.J., Brown, S.W., Meuter, M.L.: Technology Infusion in Service Encounters. Journal of the Academy of Marketing Science 28, 138–149 (2000)
33. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2, 1–135 (2008)
34. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics 35, 399–433 (2009)
35. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: International AAAI Conference on Weblogs and Social Media, 10, pp. 178–185. Menlo Park, CA, USA (2010)
36. Abbasi, A., Chen, H.: Affect Intensity Analysis of Dark Web Forums. In: Proceedings of Intelligence and Security Informatics. Chengdu, China (2017)

37. Thelwall, M., Buckley, K., Paltoglou, G., Di Cai, Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology 61, 2544–2558 (2010)
38. Higashinaka, R., Minami, Y., Dohsaka, K., Meguro, T.: Issues in Predicting User Satisfaction Transitions in Dialogues: Individual Differences, Evaluation Criteria, and Prediction Models. In: Lee, G.G., Mariani, J., Nakamura, S. (eds.) Spoken Dialogue Systems for Ambient Environments. Seond International Workshop, IWSDS 2010, Gotemba, Shizuoka, Japan, October 1-2, 2010. Proceedings, 6392, pp. 48–60. Springer, New York (2010)
39. Wang, Y., Lu, X., Tan, Y.: Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. Electronic Commerce Research and Applications 29, 1–11 (2018)
40. Gnewuch, U., Morana, S., Adam, M., Maedche, A.: Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: in Proceedings of the 26th European Conference on Information Systems (ECIS), Portsmouth, United Kingdom, June 23-28.
41. Logacheva, V., Burtsev, M., Malykh, V., Poluliakh, V., Rudnicky, A., Serban, I., Lowe, R., Prabhumoye, S., Black, A.W. and Bengio, Y.: A Dataset of Topic-Oriented Human-to-Chatbot Dialogues, http://convai.io/2017/data/dataset_description.pdf (Accessed: 30.08.2018)
42. Corredera Arbide, A., Romero, M., Moya Fernández, J.M.: Affective computing for smart operations: a survey and comparative analysis of the available tools, libraries and web services. International Journal of Innovative and Applied Research 5, 12–35 (2017)
43. Gilbert, C.H.E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International AAAI Conference on Weblogs and Social Media. Ann Arbor, MI, USA (2014)
44. Nielsen, F.Å.: A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903 (2011)
45. Cohen, J.: A power primer. Psychological bulletin 112, 155–159 (1992)
46. Hone, K.: Empathic agents to reduce user frustration. The effects of varying agent characteristics. Interacting with Computers 18, 227–245 (2006)
47. Klein, J., Moon, Y., Picard, R.W.: This computer responds to user frustration. Theory, design, and results. Interacting with Computers 14, 119–140 (2002)
48. Taylor, R.: Interpretation of the Correlation Coefficient: A Basic Review. Journal of Diagnostic Medical Sonography 6, 35–39 (1990)

Self-Tracking and Gamification: Analyzing the Interplay of Motivations, Usage and Motivation Fulfillment

Henner Gimpel[1], Niclas Nüske[1], Timon Rückel[1], Nils Urbach[2] and Matthias von Entreß-Fürsteneck[2]

[1] FIM Research Center, University of Augsburg, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Augsburg, Germany
{henner.gimpel,niclas.nueske,timon.rueckel}@fim-rc.de
[2] FIM Research Center, University of Bayreuth, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Bayreuth, Germany
{nils.urbach,matthias.vonentress}@fim-rc.de

Abstract. The usage of wearable self-tracking devices has emerged as a big trend in lifestyle and personal optimization concerning health, fitness, and well-being. In this context, gamification elements have the potential to contribute to achieving desired user behavior. However, it is not fully understood to which extent the users perceive their self-tracking motivations as being fulfilled through the usage of a wearable self-tracking device, and how gamification affects the interplay of self-tracking motivations, wearable self-tracking device usage, and motivation fulfillment. To address this research gap, we develop a conceptual model and validate it with survey research and structural equation modeling. We find that self-tracking helps users to unexpectedly fulfill motivations without previously striving for them and that significant differences exist between the gamification users and non-users with respect to their motivations by self-entertainment and self-design.

Keywords: Self-tracking, gamification, wearable self-tracking devices, motivation fulfillment, five factor framework of self-tracking motivations

1 Introduction

The engagement in self-tracking has recently emerged as a big trend in personal optimization and lifestyle [1]. Self-trackers regularly gather data about themselves – often related to their bodily functions and everyday habits – and then analyze the data to produce statistics and other analyses, such as images and diagrams [2], [3]. Devices used for this practice include for example smartphones, tablet computers, and so-called wearables. These wearable self-tracking devices benefit from sensors getting smaller as well as more compactly integrated [2]. Wearable self-tracking devices are, for example, smartwatches, wristbands, patches, clip-on devices, and jewelry or textiles with embedded sensors which measure bodily functions or physical activity [4]. The hype about self-tracking is also driven by the fact that "the new possibilities through

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

technology have opened up a world that offers new ways to get to know oneself and to gain a profound, fact-based understanding of collected self-related data" [5, p. 13].

In this regard, research on self-tracking has also emerged as a distinct stream within the IS community in recent years, studying various facets of the phenomenon [3], [6–8]. One of these facets is dedicated to understand the role of the user's motivations to engage in the practice of self-tracking. Therefore, Gimpel et al. developed a five factor framework of self-trackers' deep underlying motivations [5], while Baumgart and Wiewiorra [6] analyzed what motivations to start self-tracking drive different self-tracking activities and how different levels of self-control influence the tracking behavior of consumers and their expenditures. However, from an end-to-end perspective, a still unanswered question is to which extent the user's initial motivations are actually fulfilled through the practice of self-tracking. We therefore aim to advance this research path by investigating to which extent the users actually perceive the motivations to self-track as being fulfilled by using their wearable self-tracking devices:

RQ1*: How does the usage of wearable self-tracking devices influence the user's perceived fulfillment of the initial motivations?*

In the context of self-tracking motivation and motivation fulfillment, the practice of gamification should be considered. Gamification is a powerful method for motivating and influencing people [9]. Its term arose from the digital media industry [10] and describes the idea of using game design elements in non-game contexts [10]. One might think that gamification relates to only the motivational factor self-entertainment – below we do however argue theoretically and show empirically that gamification also significantly relates to other motivational factors. Within self-tracking experience, the application of gamified elements has the potential to change the user's behavior [11]. For example, gamification elements such as rewards, levels, leaderboards, goal-setting, and feedbacks [11], [12] are attributed to facilitate the attractiveness of monotonous physical activities [13] and therefore motivate users to become more active [12]. Consequently, when investigating self-tracking motivations and motivation fulfillments, the concept of gamification should be considered as it can be expected to influence the relationships between Gimpel et. al's [5] self-tracking motivations, actual wearable self-tracking device usage, and fulfillment of the initial motivations. Therefore, we also strive to answer the following research question:

RQ2*: How does the usage of gamification elements within the wearable self-tracking device influence the interplay of self-tracking motivations, wearable self-tracking device usage, and motivation fulfillment?*

To answer our two research questions, we develop and test a conceptual model based on the research models of Gimpel et al. [5] as well as Baumgart and Wiewiorra [6]. Further, we investigate the influence of the motivational factors of the five factor framework on the self-tracking usage and ultimately the influence of usage on the motivation fulfillment. Finally, we integrate gamification usage as a moderator to test the effect on the interplay of self-tracking motivations, wearable self-tracking device usage, and motivation fulfillment.

2 Foundations

2.1 Wearable self-tracking device usage and motivations

Wearable self-tracking devices can be assigned to the category of personal information and communication technology (ICT) devices since they are mobile (used on, e.g., the user's wrist), are adopted by individuals for their own personal usage, and enable users to engage in various activities with one device [14], [15]. To understand the users adoption of these devices, device-specific research was conducted for smartwatches [16], [17] and for fitness-trackers [18].

Further, on a more comprehensive level, Pfeiffer et al. examined what factors drive the user's pre-adoption of wearable self-tracking devices, showing perceived usefulness, perceived enjoyment, social influence, trust, personal innovativeness, and perceived support of well-being to be the major drivers for the intention to use wearable self-tracking technologies [1]. In addition, Buchwald et. al. extended research in this area by developing a model explaining post-adoption of self-tracking devices and showed that self-tracking device usage is influenced by continuance as well as discontinuance factors [19].

In contrast to these adoption models which focus on the user's perceptions about the characteristics of the self-tracking technology and its usage, Gimpel et al. developed a five factor framework of self-tracking motivations. This comprehensive study identifies and characterizes the deeper underlying motivations of users to engage in the practice of self-tracking [5]. Those five motivations are:

- **Self-entertainment:** Being motivated by the fun and ludic aspects of self-tracking. Key drivers are the enjoyments of getting lost totally in self-tracking activities, forgetting about time while doing so or playing around with numbers, statistics etc.
- **Self-association:** Being motivated by self-individualizing aspects within a community as well as the prospect of community membership. Respective reasons causing self-tracking activities are such as the urge of comparing own results to others, helping or inspiring others, and presenting oneself to them.
- **Self-design:** Being motivated by the chances of self-optimization such as the desire to control, optimize or even manipulate certain aspects of one's life or the enjoyment of being one's own master.
- **Self-discipline:** Being motivated by the self-gratification possibilities of self-tracking. Decisive aspects are the facilitation of one's self-discipline, the motivation to keep on working for a goal and the chance to reward oneself.
- **Self-healing**: Being motivated by the possibilities of self-tracking to take care of one's own health. Major factors are the aspiration of being independent from traditional medical treatments and the distrust in the healthcare system as well as classical therapies.

Gimpel et al.'s results show on the one hand that more self-tracking motivation on any of the single factors enhances the number parameters tracked as well as the time spent on self-tracking. These two constructs are defined by Gimpel at. al as self-tracking activity. On the other hand, motivation from every factor is rather independent from demographic factors (age, gender) and of personality traits (e.g. openness, conscientiousness or extraversion). Baumgart and Wiewiorra [6] further analyzed how different levels of self-control influence the tracking behavior of consumers and their expenditures for self-tracking software and hardware as well as what motivations to start self-tracking drive different self-tracking activities. They found out that the motivation of increasing one's performance as well as the number of tracked physical parameters are the key drivers of self-tracking usage frequency and accumulated expenditures. Further, customers that started self-tracking out of pure curiosity spend significantly more on self-tracking software, services and hardware and are at the same time more likely to track parameters from a wider variety of categories. Finally, they also showed that higher levels of self-control increase the odds of consumers tracking physical parameters and spending more on self-tracking software and hardware.

2.2 Gamification in the context of self-tracking

Gamification is the use of game design elements in non-game contexts [10] for changing people's behavior and driving participation as well as engagement [9]. Gamification, often interchangeably called "gamified services" [20], "gamefulness", or "gameful design" [21], also aims at the enhancement of positive patterns in service use like increasing quality and productivity of user actions, social interactions, or user activity [22].

Gamification can be reached by integrating game mechanics or elements and game dynamics. These terms are closely related and sometimes used synonymously [9]. Game elements are composed of multiple facets of "game play" [23] in the form of various actions, behaviors, and control mechanisms. While literature offers a wide range of different gamification elements [12], [23–25], rewards, levels, leaderboards, goal-setting, and feedbacks are specific gamification elements particularly considered in the context of self-tracking [11], [12]. They are the means which are used to create a compelling and appealing user experience [9] and ensure the user's engagement and his continuance in system usage [23]. Thereby, game dynamics, e.g., status, altruism, or achievement, are defined as the desires and motivations triggered by game elements. They are the universal human needs across genders, cultures, demographics, and generations which appropriate sets of game mechanics aim to satisfy [9]. Overall, gamification elements can be seen as the means which are used to satisfy game dynamics and thus, ultimately, fulfill the inherent underlying self-tracking motivations.

As mentioned before, self-trackers strive for optimizing certain aspects of their lives [5]. Especially with challenging and difficult behavior patterns for such self-optimization, users' motivation needs to be maintained in the long run. This is where gamified self-tracking applications which are designed to change the users' behavior [11], [26] come into play. One possible underlying intention could be to motivate them to become more active by making physical activity more enjoyable [12]. For example, monotonous physical activities such as running workouts can gain attractiveness by more intensively integrating the user into the application [13]. Gamification is also able to contribute positively to the usage of self-tracking apps as long-term goals can be broken down into sub-goals that can be attained more quickly. The gamification element challenges, for example, allows the user to repeatedly achieve short-term targets set by the application and might reward the user afterwards. A user planning to lose 20 kg by running might feel discouraged at first due to the long way to go. But as the application motivates the user to do single and short workouts step-by-step, the sub-goals are easier to realize. This supports the user's motivation to continuously strive for his goals [27].

Next to positive impacts of gamification on motivation in the context of self-tracking applications, also negative aspects have been identified. According to the self-determination theory of human motivation [28], competence, relatedness, and autonomy are the three innate psychological needs that determine motivation. On the one hand, intrinsic motivation gets enhanced when these needs are satisfied, but on the other hand, they diminish intrinsic motivation when they are thwarted [12]. Generally, game-play is voluntary as well as free of consequences and hence facilitates perceived autonomy, which is intrinsically motivating. But when it comes to gamified systems offering rewards or social comparison (e.g., leaderboards), their use is not necessarily voluntary or free of consequence. This might thwart perceived autonomy and hence intrinsic motivation [29]. Taken to a more general level, Nicholson [30] claims that by artificially integrating gamification elements into non-game activities, motivation will be reduced in the long run.

Besides the influence of gamification on motivations, Wellmann and Bittner [27] as well as Gal-Oz and Zuckerman [12], expanded the research stream by investigating the influence of gamification on the user's absolute, measurable goal achievement. They examined whether a gamified version of a smartphone app can affect self-tracker's physical activity. Wellmann and Bittner discovered that gamification elements within a running app can increase the user's movement behavior as their running distance was significantly larger [27]. In contrast, Gal-Oz and Zuckerman concluded that their gamified application which measures walking is only as effective as the version excluding gamification elements [12].

3 Conceptual development

3.1 Motivations and usage

IS usage can be described as the "degree and manner in which an IS is utilized by its users" [31, p. 6]. While perceptions of characteristics of an information system (e.g. perceived ease of use or usefulness) in general and self-tracking-specific influencing factors of usage have been extensively studied before (e.g., [1], [19], [32–35]), we deliberately focus on the user's underlying motivations and assume that those influence the usage behavior of a wearable self-tracking device as well. For example, the desire for self-design concerning sleep-optimization can be fulfilled by an ongoing monitoring of sleeping patterns with a sleep-tracker, thus inducing its usage. Therefore, we adapt the previously described five motivational factors identified by Gimpel et al. [5] and hypothesize:

The motivations for self-entertainment (H1.1), self-association (H1.2), self-design (H1.3), self-discipline (H1.4), and self-healing (H1.5) have a positive effect on the usage of wearable self-tracking devices.

3.2 Motivation fulfillment

After the initiation of wearable self-tracking device usage through self-tracking motivations, we assume that the continuous usage of a wearable self-tracking device leads to the perceived fulfillment of the initial motivations. For example, the initial motivation for self-discipline causes an ongoing usage of a device in terms of setting and controlling testable goals like the number of steps walked or calories burned. With the ongoing feedback of the device on these measures, the user feels his need for self-discipline being fulfilled by the device. In this regard, we define motivation fulfillment as the perceived fulfillment of the intrinsic desires reflected in the manifestation of a motivation. We further stay with five factor framework of self-tracking motivations [5], but now do consider the motivations fulfillment and hypothesize:

Wearable self-tracking device usage positively affects the user's motivation fulfillment of self-entertainment (H2.1), self-association (H2.2), self-design (H2.3), self-discipline (H2.4), and self-healing (H2.5).

3.3 Moderating effect of gamification usage

Gamification has often shown to have positive effects on motivation [9], [11–13], [27] and distinct goal achievement [27] in the context of self-tracking. A literature analysis as well as a self-conducted analysis of the top 20 iOS applications within the category of health and fitness has shown that levels, rewards, challenges, and leaderboards can be considered as the most relevant gamification elements [11], [12]. To adapt this characteristic of gamification to the context of self-tracking, we conjecture a moderating impact of gamification usage, which influences the effect of motivations for self-tracking on the actual wearable self-tracking device usage. Consequently, we suppose the positive effects of gamification on motivation to be predominant and posit:

Gamification usage positively moderates the effect of the motivations self-entertainment (H3.1.1), self-association (H3.1.2), self-design (H3.1.3), self-discipline (H3.1.4), and self-healing (H3.1.5) on wearable self-tracking device usage.

As gamification can also, in general, enhance system usage [13], we adapt this characteristic of gamification to the self-tracking context. We assume that the user's continuous usage of wearable self-tracking devices leads to a satisfaction of her or his motivations and therefore hypothesize gamification usage to also moderate the effect of wearable self-tracking device usage on motivation fulfillment. Again, we suppose the positive effects of gamification on motivation fulfillment to be predominant and hypothesize:

Gamification usage positively moderates the effect of wearable self-tracking device usage on the motivation fulfillment of self-entertainment (H3.2.1), self-association (H3.2.2), self-design (H3.2.3), self-discipline (H3.2.4), and self-healing (H3.2.5).

4 Survey design and procedures

We chose a quantitative-empirical research approach to validate our conceptual research model because it allows for a statistical generalization on the basis of results which are representative of the whole population at a lower cost than collecting the data for the whole population [36]. To this end, we crafted a survey instrument. We began this process by using, wherever possible, established and validated measurement scales and adapted them if necessary to ensure that the focus of our study is centrally reflected in each of the statements. Each of the item statements was measured with a seven-point Likert scale [37]. All constructs are measured reflectively.

To further enhance the survey instrument's comprehensibility and validity, we conducted a pretest with six researchers and incorporated their qualitative feedback. Ultimately, we used our survey instrument to collect empirical data via an online-survey tool.

4.1 Construct operationalization

We measure both self-tracking motivation (M) and motivation fulfillment (F) based on the five factors self-entertainment (SE), self-association (SA), self-design (SDe), self-discipline (SDi), and self-healing (SH) [5]. We utilize all items from [5] to measure both the current self-tracking motivation and motivation fulfillment. For the measurement of current self-tracking motivation, the items represent answers to the originally proposed question "I am self-tracking because…" (Table 1 lists all items) and range from "strongly disagree" to "strongly agree". For each item, this question regarding self-tracking motivation was immediately followed by an evaluation of the phrase "I actually fulfill this goal by self-tracking." to capture motivation fulfillment. The answer-options range from "not fulfilled as I expected" to "fulfilled way more than I expected". In addition, we added the scale item "not applicable as not a goal of mine" in the motivation fulfillment.

Table 1. Operationalization of self-tracking motivation [5]

Constructs	Items
	I am self-tracking because…
Self-entertainment (SE)	... I enjoy getting lost totally in self-tracking activities.
	... I like playing around with numbers/statistics etc.
	... I like playing around with my smartphone/technical device etc.
	... I enjoy forgetting about time while doing so.
	... it is fun and entertaining.
Self-association (SA)	... I want to help/inspire others.
	... the way I'm doing it is interesting for others/might help others.
	... I want to compare my results to others.
	... I want to present myself to others.
Self-design (SDe)	... I want to control what I'm doing with my life.
	... I try to manipulate certain aspects in my life.

	... I enjoy being my own master.
	... I'm interested in how certain things in (my) life interact.
	... it helps me to optimize the way I'm living.
Self-discipline (SDi)	... it motivates me to keep on working for a goal.
	... It allows me to reward myself.
	... it facilitates my self-discipline.
Self-healing (SH)	... I don't trust in the healthcare system/classic therapies.
	... I want to be independent from traditional medical treatments.

Further, we self-developed two measurement items for wearable self-tracking device usage (WSTDU) based on Burton-Jones and Straub [38], Davis et. al [39] as well as Venkatesh and Davis [40]. The answer-options range from "Less than few times a month" to "Almost 24 hours a day". Regarding gamification usage, we differentiate between active self-tracking users if they at least use one of the four considered gamification elements rewards, levels, leaderboards, and challenges and those who do not engage with any of these elements. Thus, gamification usage represents a binary variable. The final operationalization of wearable self-tracking device usage (WSTDU) and gamification usage is shown in Table 2.

Table 2: Operationalization of wearable self-tracking device usage (WSTDU)

Constructs	**Items**
WSTDU	On average, how frequently do you (passively) collect data with your wearable self-tracking device?
	On average, how frequently do you actively engage with your wearable self-tracking device (e.g., for data analysis)?
Gamification usage	Do you use the gamification element *Rewards / Levels / Leaderboards / Challenges* within your wearable self-tracking device? [Four items, one each for the four gamification elements]

4.2 Data collection

We collected data by administering our survey instrument to current active users of wearable self-tracking devices. This means that it was a prerequisite that the users actively use their device to track their fitness, health, or well-being to increase the validity of the responses. Users who do not yet use or have already stopped using their devices were excluded from the survey. We explained to the participants the concept of self-tracking and the function of the different gamification elements to receive more valid responses. To gather our data, we offered English and German versions and distributed the invitation message to participate in our study in online social networks (e.g., Facebook), online business networks (e.g., Xing and LinkedIn), instant-messaging services (e.g., WhatsApp), and the e-learning system of one of the authors' universities. We decided to openly circulate our invitation to allow for a snowball effect within social media. Overall, we received 359 responses. We excluded non-self-tracking users and incomplete answers (270 in total) which left us with 89 remaining responses. Of these 89 respondents, 53% indicated that they are actively using at least one of the four gamification elements. The average time of usage for the wearable self-tracking devices was 20 months. 84% use smartphone apps for self-tracking, 33% an activity tracker, 17% a smartwatch and 10% another form of device or application. On a seven-point Likert scale ranging from light user (1) to heavy user (7), 52% of the sample group consider themselves as medium self-tracking user type (4) or higher. On a seven-point Likert scale ranging from strongly disagree (1) to strongly agree (7), 64% either agree (6) or strongly agree (7) to be interested in trying out new technical devices. 57% agree or strongly agree that they actively take care of their health and well-being and 51% that they see themselves as sportive.

5 Data analysis and results

We tested measurement properties and hypotheses with a partial least squares structural equation modeling approach (PLS-SEM) and multi-group analysis (MGA) [41], [42] using the software SmartPLS Version 3.2.6 [43]. Even though PLS-SEM has its limitations [44], we chose it as an established approach in the IS research discipline and for our study especially due to the relatively small sample size [45], [46].

5.1 Measurement model

Concerning outer loadings, we set the critical threshold at 0.70 [47]. The outer loadings of the fourth item of self-entertainment motivation and motivation fulfillment, the second item of self-discipline motivation and motivation fulfillment, and the third item of self-design motivation and motivation fulfillment are lower than 0.70. We excluded them from our measurement model. The first two items of self-association motivation and the last two items of self-association motivation fulfillment exhibit lower outer loadings than 0.70 as well. Due to the nature of our measurement model, dropping these items would lead to an asymmetric inconsistency between the constructs. We therefore further examined the data and the operationalization of the construct. The results suggest that the operationalization may describe two different facets of self-association, one more directed towards altruism, the other more towards self-presentation. Hence, we decided to not further consider the results of self-association. Furthermore, the first item of self-entertainment motivation and the first and fifth item of self-entertainment motivation fulfillment do not reach the critical threshold of 0.70. But as they still exceed 0.60, which is deemed high [48], we considered them as marginal and did not exclude them from our measurement model. All other items, including active as well as passive use frequency of the construct wearable self-tracking device usage are greater than the critical threshold. Adhering to standard validation guidelines [49–51], we tested the reflective measurement model in terms of internal consistency reliability, indicator reliability, convergent validity, and discriminant validity. The internal consistency reliabilities (composite reliability) of multi-item scales modeled with reflective indicators is 0.81 or greater, suggesting that scales were reliable. In addition, the Cronbach's Alpha values are, except for self-association, 0.70 or greater, hence showing a good internal consistency of our scale. The average variance extracted is consistently greater than the critical threshold of 0.50. Hence, we conclude that convergent validity has been established. Further, to check for discriminant validity, we applied the Fornell-Larcker Criterion as a conservative measure [52]. The square root of each construct's AVE is greater than its highest correlation with any other construct, hence discriminant validity has been established, too.

5.2 Structural model

To assess the significance levels of our structural model including the MGA, we applied bootstrapping with 5,000 sub-samples (no sign changes). Table 3 presents the results for the entire group and for the sub-groups of gamification users and non-users. Relating to the 20 hypotheses posed, 4 could not be tested due to measurement problems with self-association. Of the remaining 16 hypotheses, 7 are supported by the data. These seven hypotheses are discussed in the following. In that, we apply a 10% significance level which appears reasonable given the relatively small sample size, especially in the subgroups. Our data support that the motivation for self-entertainment increases the wearable self-tracking device usage and the latter positively influences the user's self-entertainment motivation fulfillment (H1.1 and H2.1). Further, the multi-group analysis of gamification users and non-users shows a significant difference between the two groups, with a significantly higher effect of motivation for self-entertainment on usage within the group of gamification users (H3.1.1).

Table 3. PLS-MGA results

Hypothesis	Complete		Non-gami-fication users	Gamification users	Group delta
	n = 89		n = 42	n = 47	
	Path coefficients	R^2	Path coefficients		
M-SE → WSTDU	0.276 **	0.163	0.092	0.428 **	0.337 +
M-SDe → WSTDU	0.195		0.423 +	0.041	0.382 +
M-SDi → WSTDU	0.115		0.027 +	0.137	0.110
M-SH → WSTDU	-0.076		-0.248	0.044	0.292 +
WSTDU → F-SE	0.242 *	0.058	0.238	0.297	0.059
WSTDU → F-SDe	0.326 ***	0.106	0.491 ***	0.235	0.256
WSTDU → F-SDi	0.321 ***	0.103	0.434 ***	0.218	0.215 +
WSTDU → F-SH	0.139	0.019	0.244 *	0.168	0.076
Significance levels: + 10%, * 5%, ** 1% *** 0.1% | n = number of cases					

Self-design shows significant results as well: Wearable self-tracking device usage significantly increases the user's self-design motivation fulfillment (H2.3). Additionally, within the multi-group analysis, the influence of motivation for self-design on usage is significantly higher in the group of non-gamification users (H3.1.3).

Also, our results reveal that wearable self-tracking device usage significantly increases the self-discipline motivation fulfillment (H2.4). Finally, the multi-group analysis results show that the influence of wearable self-tracking device usage on the self-discipline motivation fulfillment is significantly higher in the non-gamification group (H3.2.4).

6 Discussion

Taking a comprehensive look at our results, we acknowledge the relatively low R^2 values of the dependent variables. However, the results are reasonable since our study specifically only aims on the user's deeper underlying motivations of self-tracking and does not take the user's perceptions about the characteristics of the self-tracking technology and its usage into account which were analyzed in other dedicated acceptance studies [1], [19]. Looking further into the details of our results, self-entertainment is the key motivation to engage in the practice of self-tracking as it is the only effect on wearable self-tracking device usage that is significant. Users seem to be driven by the entertainment possibilities which allow them to experience fun and play around with their collected data and statistics. Concerning the multi-group analysis, the effect is even more pronounced among gamification users and significantly differs from that of non-gamification users. This observation confirms that the playful elements of gamification reinforce the urge to self-track due to ludic motivation.

In contrast, the motivations self-design, self-discipline and self-healing are not found to drive wearable self-tracking device usage per se. However, the MGA shows that the motivation for self-design has a significantly higher influence on usage for non-gamification users. A potential reason could be, that non-gamification users who pursue control and optimization engage in these activities with a more serious mindset, thus deliberately ignore playful gamification elements because they might not support or even distract them.

Moving on to the relationships between wearable self-tracking device usage and motivation fulfillment, results show that usage significantly increases the perception that the preexisting desire for self-entertainment is fulfilled. The users of wearable self-tracking devices feel that their wishes to entertain themselves are met in the process of self-tracking. For self-design and self-discipline, however, we observe significant positive effects of usage on perceived motivation fulfillment without significant preexisting connections between their motivation and usage. Hence, users might not necessarily start self-tracking due to a striving for self-design or self-discipline. Nevertheless, as soon as they are active wearable self-tracking device users, they seem to realize positive effects such as being able to take control of and optimize their lives, gaining knowledge about interactions of certain things within their lives (self-design), facilitating their self-discipline, or being motivated to keep on working on goals (self-discipline). A further look at the group of non-gamification users reveals that they clearly and highly significantly perceive their motivation for self-discipline as better fulfilled than gamification users. This fact seems counterintuitive; however, a possible explanation here might also be that gamification elements do not support motivation fulfillment but rather distract the users from it.

7 Conclusion

The purpose of this study was to create a basis for future research regarding the analysis of the interplay of self-tracking motivations, usage and motivation fulfillment. Therefore, our paper investigates how Gimpel et. al's [5] motivational factors for self-tracking influence the actual usage of wearable self-tracking devices, to which extent the users actually perceive these motivations as being fulfilled in the process of using them, and how gamification affects this interplay of self-tracking motivations, wearable self-tracking device usage, and motivation fulfillment. We found the motivation for self-entertainment to represent the crucial driver of wearable self-tracking device usage and ultimately usage as important driver for the motivation fulfillment of the three factors self-entertainment, self-discipline, and self-design. Further, both the motivation as well as the motivation fulfillment are moderated by gamification usage. Gamification users are more motivated by self-entertainment, non-gamification users more by self-design. In addition, non-gamification users tend to have higher levels of motivation fulfillment, except for self-entertainment. Hence, in designing self-tracking devices and apps and potentially integrating gamification elements, one should carefully consider the diverse effects of gamification.

Our study has three main limitations: First, as common in research on motivation, survey responses are self-reports. Second, our results are based on a relatively small sample size of 89 respondents which may distort the results. Future research on this topic should be built on a broader database which enables more precise and refined results. Additionally, multiple surveys at different points in time would enable empirically validated statements on continuous usage. Lastly, the influence of gamification is only explained based on the distinction between gamification users and non-users. For future research, the differentiation between the four major gamification elements would allow for more detailed insights of the influence of gamification use. Additionally, coming research could combine the research on self-tracking motivations with the research on the user's perceptions about the characteristics of the self-tracking technology which might further increase the understanding of the phenomenon.

Generally, our research contributes to the domain of self-tracking and gamification as it advances the understanding how the usage of wearable self-tracking devices influences the user's perceived fulfillment of the initial motivations, and how gamification elements affect this interplay. Thereby, we found evidence that next to the motivation of increasing one's performance [6], striving for self-entertainment is a key driver for using wearable self-tracking devices, and that the usage ultimately increases the perceived fulfillment of the user's motivations for self-entertainment, self-discipline as well as self-design. Furthermore, gamification elements might not support motivation fulfillment but rather distract users of wearable self-tracking devices from it. Our findings have three additional main practical implications: First, potential users of wearable self-tracking devices should be aware that self-tracking might help them to fulfill motivations which they have not previously been aware of. Second, we suggest that designers and manufacturers of wearable self-tracking devices consider addressing the entirety of motivational factors. This might improve their product attractiveness and let them reach more customers. Lastly, the use of gamification elements should be up to the user as their mandatory usage might not always support usage and motivation fulfillment.

References

1. Pfeiffer, J., von Entreß-Fürsteneck, M., Urbach, N., Buchwald, A.: Quantify-Me. Consumer Acceptance of Wearable Self-Tracking Devices. In: Proceedings of the 24th European Conference on Information Systems (2016)
2. Choe, E.K., Lee, N.B., Lee, B., Pratt, W., Kientz, J.A.: Understanding Quantified-Selfers´ Practices in Collecting and Exploring Personal Data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'14, pp. 1143-1152 (2014)
3. Sjöklint, M., Constantiou, I., Trier, M.: The Complexities of Self-Tracking - An Inquiry into User Reactions and Goal Attainment. In: Proceedings of the 23rd European Conference on Information Systems (2015)
4. Swan, M.: Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. JSAN vol. 1, iss. 3, pp. 217–253 (2012)
5. Gimpel H., Nißen M., Görlitz R. A.: Quantifying the Quantified Self. A Study on the Motivation of Patients to Track Their Own Health. In: Proceedings of the 34th International Conference on Information Systems (2013)
6. Baumgart, R., Wiewiorra, L.: The Role of Self-Control in Self-Tracking. In: Proceedings of the 37th International Conference on Information Systems (2016)
7. Buchwald, A., Letner, A., Urbach, N., von Entreß-Fürsteneck, M.: Towards Explaining the Use of Self-Tracking Devices. Conceptual Development of a Continuance and Discontinuance Model. In: Proceedings of the 36th International Conference on Information Systems (2015)
8. Lupton, D.: Self-tracking cultures. Towards a sociology of personal informatics. In: Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design, pp. 77–86 (2014)
9. Bunchball Inc.: Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior. Bunchball Inc. (2010)
10. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From Game Design Elements to Gamefulness: Defining “Gamification”. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15. ACM Press, New York, USA (2011)
11. Lister, C., West, H.J., Cannon, B., Sax, T., Brodegard, D.: Just a Fad? Gamification in Health and Fitness Apps. JMIR Serious Games vol. 2, iss. 2, e9 (2014)
12. Zuckerman, O., Gal-Oz, A.: Deconstructing gamification. evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Personal and Ubiquitous Computing vol. 18, iss. 7, pp. 1705–1719 (2014)
13. Rapp, A., Marcengo, A., Console, L., Simeoni, R.: Playing in the wild. Enhancing user engagement in field evaluation methods. In: Proceedings of the 16th International Academic MindTrek Conference, p. 227 (2012)

14. Hong, S.-J., Tam, K.Y.: Understanding the Adoption of Multipurpose Information Appliances: The Case of Mobile Data Services. Information Systems Research vol. 17, iss. 2, pp. 162–179 (2006)
15. Scheepers, R., Middleton, C.: Personal ICT Ensembles and Ubiquitous Information Systems Environments: Key Issues and Research Implications. Communications of the Association for Information Systems vol. 33, iss. 1 (2013)
16. Kim, K.J., Shin, D.-H.: An acceptance model for smart watches. Implications for the adoption of future wearable technology. Internet Research iss. 25, vol. 4, pp. 527–541 (2015)
17. Chuah, S.H.-W., Rauschnabel, P.A., Krey, N., Nguyen, B., Ramayah, T., Lade, S.: Wearable technologies. The role of usefulness and visibility in smartwatch adoption. Computers in Human Behavior 65, 276–284 (2016)
18. Sol, R., Baras, K.: Assessment of activity trackers. In: Lukowicz, P., Krüger, A., Bulling, A., Lim, Y.-K., Patel, S.N. (eds.) Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 570–575. ACM, New York, NY (2016)
19. Buchwald, A., Letner, A., Urbach, N., von Entreß-Fürsteneck, M.: Insights into Personal ICT Use: Understanding Continuance and Discontinuance of Wearable Self-Tracking Devices. In: Proceedings of the 26th European Conference on Information Systems (2018)
20. Hamari, J., Huotari, K.: Defining Gamification - A Service Marketing Perspective. In: Proceedings of the 16th International Academic MindTrek Conference, pp. 17–22 (2012)
21. Stieglitz, S., Lattemann, C., Robra-Bissantz, S., Zarnekow, R. (eds.): Gamification. Using Game Elements in Serious Contexts. Springer, Cham, Switzerland (2017)
22. Hamari, J., Koivisto, J., Sarsa, H.: Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In: 47th Hawaii International Conference on System Science (HICSS) (2014)
23. Burke, M., Hiltbrand, T.: How Gamification Will Change Business Intelligence. Business Intelligence Journal vol. 16, iss. 2, pp. 8–16 (2011)
24. Zichermann, G., Cunningham, C.: Gamification by design. Implementing game mechanics in web and mobile apps. O'Reilly, Sebastopol, CA (2011)
25. Thiebes S., Lins S., Basten D.: Gamifying Information Systems - A Synthesis of Gamification Mechanics and Dynamics. In: Proceedings of the 22nd European Conference on Information Systems (2014)
26. Hamari, J., Koivisto, J.: Working out for likes: An empirical study on social influence in exercise gamification. Computers in Human Behavior vol. 50, pp. 333–347 (2015)
27. Wellmann, C., Bittner, J.V.: Gamification-Elemente bei Apps zur Bewegungsförderung. In: Pabst Science Publishers (ed.) Wirtschaftspsychologie. E-Interventionen auf individueller und Gruppenebene, 4, pp. 28–39 (2016)
28. Deci, E.L., Ryan, R.M.: Intrinsic motivation and self-determination in human behavior. Plenum Press, New York (1996)
29. Deterding, S.: Situated motivational affordances of game elements: A conceptual model. In: Tan, D. (ed.) Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems. ACM, New York, NY (2011)
30. Nicholson, S.: A User-Centered Theoretical Framework for Meaningful Gamification. In: Steinkuehler, C. (ed.) Games+Learning+Society 8.0 (2012)
31. Urbach, N., Müller, B.: The Updated DeLone and McLean Model of Information Systems Success. In: Dwivedi, Y.K., Wade, M.R., Schneberger, S.L. (eds.) Information systems theory, 28, pp. 1–18. Springer, New York, London (2011)
32. Venkatesh, Morris, Davis: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly vol. 27, iss. 3, p. 425 (2003)

33. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly vol. 36, iss. 1, pp. 157–178 (2012)
34. Buchwald, A., Letner, A., Urbach, N., von Entreß-Fürsteneck, M.: Towards Explaining the Willingness to Disclose Personal Self-Tracking Data to Service Providers. In: Proceedings of the 25th European Conference on Information Systems (2017)
35. Davis, F.D.: A Technology Acceptance Model for Empirically Testing New End-User Information Systems - Theory and Results. Massachusetts (1985)
36. Saunders, M., Lewis, P., Thornhill, A.: Research Methods for Business Students. Financial Times Prentice Hall, Harlow (2009)
37. Likert, R.: A Technique for the Measurement of Attidtudes. In: Woodworth (ed.) Archives of Psychology, 22, pp. 5–55. New York (1932)
38. Burton-Jones, A., Straub, D.W.: Reconceptualizing System Usage: An Approach and Empirical Test. Information Systems Research vol. 17, iss. 3, pp. 228–246 (2006)
39. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science vol. 35, iss. 8, pp. 982–1003 (1989)
40. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science vol. 46, iss. 2, pp. 186–204 (2000)
41. Wold, H.: Partial Least Squares. In: Kotz, S., Read, C.B., Balakrishnan, N., Vidakovic, B., Johnson, N.L. (eds.) Encyclopedia of statistical sciences, 53, p. 525. Wiley-InterScience, Hoboken, NJ (2005)
42. Chin, W.W.: The partial least squares approach for structural equation modeling. In: Marcoulides, G.A. (ed.) Methodology for business and management. Modern methods for business research, pp. 295-336. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (1998)
43. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS. SmartPLS GmbH, Bönningstedt (2015)
44. Tomarken, A.J., Waller, N.G.: Structural equation modeling: strengths, limitations, and misconceptions. Annual review of clinical psychology 1, 31–65 (2005)
45. Gefen, D., Rigdon, E.E., Straub, D.: An Update and Extension to SEM Guidelines for Administrative and Social Science Research. MIS Quarterly 35, 3–14 (2011)
46. Hair, J., Hollingsworth, C.L., Randolph, A.B., Chong, A.Y.L.: An updated and expanded assessment of PLS-SEM in information systems research. Industrial Management & Data Systems 117, 442–458 (2017)
47. Hair, J.F.: A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publications, Inc, Thousand Oaks (Calif.) (op. 2017)
48. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. MIS Quarterly vol. 27, iss. 1, pp. 51–90 (2003)
49. Lewis, B.R., Templeton, G.F., Byrd, T.A.: A methodology for construct development in MIS research. European Journal of Information Systems vol. 14, iss. 4, pp. 388–400 (2005)
50. Straub, D., Boudreau, M.-C., Gefen, D.: Validation Guidelines for IS Positivist Research. Communications of the Association for Information Systems vol. 13, iss. 1 (2004)
51. Urbach, N., Ahlemann, F.: Structural Equation Modeling in Information Systems Research Using Partial Least Squares. In: Association for Information Systems (ed.) Journal of Information Technology Theory and Application, pp. 5–40 (2010)
52. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research vol. 18, iss. 1, p. 39 (1981)

Erfolgsfaktoren von Augmented-Reality-Applikationen: Analyse von Nutzerrezensionen mit dem Review-Mining-Verfahren

Laura Sophie Gravemeier[1], Lisa Berkemeier[1], Oliver Thomas[1]

[1] Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Osnabrück, Deutschland
{lgravemeier,lisa.berkemeier,oliver.thomas}@uni-osnabrueck.de

Abstract. Nachdem die Leistungsentwicklung von mobilen Endgeräten der letzten Jahre die Nutzung von Augmented-Reality-Anwendungen dem Massenmarkt zugänglich gemacht hat, sind diese nun gängiger Bestandteil der App Stores. Es besteht jedoch eine Forschungslücke im Hinblick auf klare Gestaltungsrichtlinien für Augmented-Reality-Applikationen. Rezensionen in den App Stores bieten an dieser Stelle eine wertvolle Einsicht in das Nutzererlebnis. Zur Identifikation von Erfolgsfaktoren wird ein Review Mining mit den Nutzerkommentaren zu 20 Augmented Reality Applikationen durchgeführt. Durch eine Sentiment Analysis werden positive und negative Nutzererfahrungen voneinander getrennt, um darin wiederholt auftretende Lob- und Kritikpunkte durch ein Topic Modeling zu erkennen. Die Erkenntnisse erweitern die Wissensbasis durch die Identifikation von Adoptionstreibern und Entwicklungsfeldern für die Gestaltung von Augmented-Reality-Applikationen. Insbesondere technische Zuverlässigkeit und Kompatibilitätsprobleme wurden als Entwicklungsbedarf identifiziert, während die volle Ausnutzung des Potenzials des Augmented Reality Interface Designs sich als klarer Treiber zeigt.

Keywords: Augmented Reality, User Experience, Review Mining, Sentiment Analysis, Topic Modeling

1 Einleitung

Applikationen (Apps) können das Konzept Augmented Reality (AR) auf verschiedenen Endgeräten umsetzen, indem sie digitale Objekte in die reale, wahrgenommene Umgebung des Nutzers integrieren. Nach der erfolgreichen Einführung der App „Pokémon Go" werden nun verschiedene AR-Apps entwickelt und konsumiert. Neben der Anwendung in der Spieleindustrie werden AR-Apps auch in klassischen Branchen wie der Möbelindustrie und im Bildungssektor entwickelt. Hierzu stellte Apple Inc. [1] auf der World Wide Developer Conference im Juni 2017 das ARKit vor, im September antwortete Google LLC [2] im Rahmen des Google Developers Day mit der Ankündigung des ARCore. Sie dienen als Entwicklungswerkzeuge für AR-Apps und bereiten den Weg für die weitere Verbreitung von Mobile-Augmented-Reality-Apps (MAR-Apps) in den gängigen App Stores.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Obwohl die ersten Überlegungen zu Augmented bzw. Virtual Reality bereits lange zurückliegen [3], weitet sich damit erst jetzt ihre Verwendung auf eine Vielzahl von Endnutzern aus. Folglich befindet sich die wissenschaftliche Forschung zu Design und Konzeption von MAR-Apps mit Berücksichtigung des Nutzerverhaltens und deren Umsetzung derzeit noch in ihren Anfängen [4–6]. Allerdings ist die Erhebung von Nutzerdaten durch Usability- und User-Experience-Experimente mit hohen Kosten verbunden und bisher vorliegende Untersuchungen weisen häufig noch Qualitätsdefizite auf [7]. Für herkömmliche Apps stehen zwar bereits Richtlinien zur Interfacegestaltung auf Basis wissenschaftlicher Forschung zur Verfügung (z.B. [8]), allerdings können diese aufgrund der besonderen Interaktionsmöglichkeiten nur begrenzt auf AR-Apps übertragen werden.

Die Technologie eröffnet neue Möglichkeiten zur Dienstleistungserbringung seitens der Industrie und schafft neue Nutzungserlebnisse für den Endkunden. Diese Erfahrungen werden in Form von Bewertungen auf entsprechenden Plattformen, wie den App-Stores von Apple und Google, rezensiert. Diese unstrukturierten Daten können wertvolle Erkenntnisse für die Entwicklung von AR-Apps beinhalten. Ziel dieses Beitrags ist die Identifikation von Erfolgsfaktoren, aufgegliedert in Adoptionstreiber bzw. Entwicklungsfelder, die aus Rezensionen mittels Text Mining abgeleitet und analysiert werden. Die Untersuchung wird strukturiert durch die Forschungsfrage (FF) *Welche Aspekte von AR-Apps werden wiederholt positiv bzw. negativ zurückgemeldet und können daher als Erfolgsfaktoren von AR-Apps gewertet werden?* Dazu werden in Kapitel zwei zunächst die theoretischen Grundlagen der AR und des Review Mining (RM) als zentrale Konzepte erläutert. Anschließend wird in Kapitel drei die Durchführung und verwendete Methoden des RM beschrieben, um eine Rekonstruierbarkeit zu gewährleisten. In Kapitel vier werden die erzielten Ergebnisse umfassend dargestellt und in Kapitel fünf diskutiert. Abschließend wird in Kapitel sechs ein Fazit gezogen und ein Ausblick auf weiteren Forschungsbedarf gegeben.

2 Theoretische Grundlagen

2.1 Anwendungsgebiete und Interfaces von Augmented Reality Applikationen

AR beschreibt die Integration virtueller Elemente in die reale Umwelt. Der nahtlose Übergang virtueller Objekte in die reale Umgebung wird als Erweiterung der Realität verstanden. Dabei wird eine vermischte 3D Realität erzeugt, mit der in Echtzeit interagiert werden kann [9, 11]. MAR, d.h. die Verwendung von AR auf mobilen Geräten [12], wird von Smartphones geprägt [5], [13]. Smartphones stehen einer breiten Masse an Anwendern zur Verfügung und können im Gegensatz zu spezifischen Endgeräten, wie AR-Brillen, die Weitergabe der AR-Technologie von einer forschungszentrierten Entwicklung hin zum Endnutzermarkt beschleunigen. AR-Apps werden in der Regel nach Anwendungsfeldern kategorisiert [vgl. 4]. Über verschiedene Klassifikationen hinweg finden sich AR-Apps mit Bezug zu Bildung, Gesundheit, Marketing, Bauwesen und Fertigung, Gaming und Entertainment, Militär, Visualisierung, sowie Navigation und Tourismus [4], [9], [14–17].

Im Rahmen dieser Anwendungsgebiete werden AR-Apps sich nur durchsetzen können, wenn sie im Vergleich mit herkömmlichen Apps Vorteile bieten. Dabei nutzt AR das Potential der Immersion, um das Nutzererlebnis zu verbessern. So wird z.B. beim Online-Shopping mit der IKEA AR-App die Anordnung virtueller Produkte im tatsächlichen Raum und somit eine bessere Entscheidungssituation ermöglicht. Die vielfältigen Interaktionsmöglichkeiten der AR-Technologie ermöglichen neue User Interfaces (UI). Insbesondere tragbare MAR-Geräte können als „magic lens" [17] bezeichnet werden, durch die der Benutzer die Realität erweitert wahrnimmt. Die Interaktion mit den darin sichtbaren virtuellen Objekten kann in einem tangible UI [18] direkt durch die Manipulation der korrespondierenden realen Objekte erfolgen. Zusätzlich dient ein kollaboratives UI mehreren Benutzern als Basis für einen ausgeführten Prozess, sodass gemeinsam nicht nur mit realen, sondern auch mit virtuellen Objekten interagiert werden kann [19], [20]. Generell wird die Zusammenführung verschiedener Interaktionsmöglichkeiten und somit ein hybrides UI angestrebt, um eine möglichst realitätsnahe Erfahrung zu erzeugen [17]. Um die Schnittstelle zwischen Anwender und AR-System noch lebendiger zu gestalten, kommen multimodale UIs zum Einsatz. Hier werden mehrere Sinne des Benutzers für die Verarbeitung des In- und Outputs des AR-Systems verwendet [17].

2.2 Review Mining

Die Analyse von Reviews zur Zusammenfassung verschiedener Produkteigenschaften und Nutzereindrücke wird abgeleitet vom Text Mining als RM bezeichnet [21] und als Tool zur Analyse von Apps eingesetzt [22], [23]. Durch die Betrachtung von Einschätzungen der Nutzer stehen vor allem die Methoden des Opinion Mining bzw. der Sentiment Analysis (SA) im Vordergrund. Die Abgrenzung der Begriffe Opinion Mining und SA ist umstritten [24], allerdings dienen beide der Analyse von Texten mit subjektiven Inhalten im Gegensatz zu klassischen Text Mining Anwendungen [21], [25].

Eine zentrale Komponente der Analyse subjektiver Daten ist die Einschätzung ihrer Polarität bzw. der Richtung, der darin ausgedrückten Empfindungen. Diese Klassifikation kann durch verschiedene Methoden erreicht werden: mithilfe von Lexika, in denen Sentiment-Werte für enthaltene Wörter bereits definiert sind, aber auch gestützt durch maschinelle Lernverfahren wie das Natural Language Processing [25]. Diese Lernverfahren benötigen bereits klassifizierte Datensätze, um davon eine systematische Einteilung abzuleiten. Um möglichst genaue Ergebnisse zu erzielen, werden domänenspezifische Trainingsdaten benötigt. Allerdings ist die manuelle Generierung individueller Datensätze für jedes neue Anwendungsgebiet mit einem hohen Aufwand verbunden und daher werden auch domänenübergreifende Ansätze verfolgt [26]. Meist wird die Polaritätseinschätzung auf Basis des gesamten Textdokuments durchgeführt, es kann jedoch auch auf Ebene von Sätzen oder sogar Aspekten differenziert werden. Die erhöhte Granularität ermöglicht auf der einen Seite eine klarere Zuordnung der subjektiven Einschätzung zum Meinungsgegenstand, kann aber durch den Verlust von Kontext auch unerwünschte Effekte hervorrufen [27].

Zusätzlich zur Polarität der ausgedrückten Meinungen, muss auch der inhaltliche Gegenstand der Reviews ausgewertet werden. Um wiederkehrende Themen und ihre Zusammenhänge zu kategorisieren, werden Topic Modeling Methoden genutzt [28], [29]. Häufig wird Latent Dirichlet Allocation (LDA) angewandt, das übergeordnete Themen und assoziierte Inhalte in Textquellen identifiziert [30], [31]. Klassische LDA Verfahren fordern dabei eine Angabe der Topicanzahl, während das hierarchische LDA diese dynamisch selbst erlernt [32].

3 Analyse von Rezensionen über Augmented Reality Applikationen mit dem Review-Mining-Verfahren

Im Rahmen dieser Arbeit wird ein RM mithilfe der oben erläuterten Methoden durchgeführt. Hier wird zunächst der konkrete Untersuchungsablauf beschrieben, der zum Überblick in Abbildung 1 dargestellt ist. Zuerst wurden die zu untersuchenden Daten ausgewählt und extrahiert. Mit dem Ziel möglichst viele Reviews zu erfassen, wurden Daten aus den Apple App Stores aller Nationen mit Englisch als Muttersprache berücksichtigt: die Vereinigten Staaten, Kanada, das Vereinigte Königreich, Irland, Australien, Neuseeland und Südafrika [33]. Als Indikator für eine möglichst hohe Reviewdichte wurden die 20 Apps mit der höchsten Anzahl an Benutzerbewertungen im App Store der Vereinigten Staaten ausgewählt. Daten zu Download- oder Reviewanzahl können im App Store nicht eingesehen werden. Die Extraktion der Daten wurde mittels eines PHP-Skripts (modifiziert nach [34]) durchgeführt, das Datum, Rating, Titel, Inhalt und weitere Metadaten des Reviews gescrapt hat.

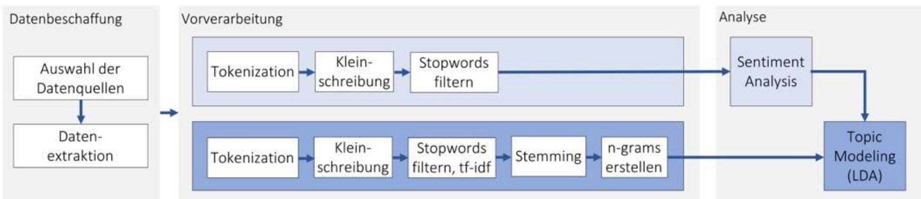


Abbildung 1. Ablauf des durchgeführten RM

Zur Analyse der Daten wurde das Tool RapidMiner Studio (Version 8.2.000) verwendet. Im ersten Schritt wurden durch das Scraping oder Spam entstandene Duplikate entfernt, wonach mit 130.262 Datensätzen gearbeitet werden konnte. Deskriptive Daten zur initialen Stichprobe sind in Tabelle 2 im Anhang dargestellt.

Als Grundlage der SA dienen die textuellen Bestandteile der gescrapten Daten, d.h. Reviewtitel und -inhalt. Diese wurden in ihre einzelnen Wortbestandteile aufgeteilt (Tokenization) und in Kleinbuchstaben umgewandelt. Zusätzlich wurden sogenannte Stopwords herausgefiltert, d.h. in der englischen Sprache grundsätzlich gehäufte Worte oder auf den App-Inhalt bezogene Worte ohne Relevanz für die ausgedrückte Meinung. Anschließend wurde auf Basis der verbliebenen Tokens eine SA durchgeführt. Da kein domänenspezifisches Training Set zur Verfügung stand, wurde ein Sentiment Score mit dem domänenübergreifenden Tool SentiWordNet 3.0 [35] berechnet.

Vorbereitend für das Topic Modeling Verfahren wurde der Datensatz zunächst verkleinert, um eine ausgewogene Zusammensetzung der Daten zu erreichen. Zuvor lag der Reviewanteil der drei führenden Apps bei 91,18% und Pokémon Go dominierte mit 86.217 von 130.262 Dateneinträgen. Um einen zu starken Einfluss dieser drei Apps zu verhindern und gleichzeitig die Stichprobe nicht unnötig zu verkleinern, wurden per Zufallsauswahl jeweils nur 5.000 ihrer Reviews zur Untersuchung ausgewählt und damit insgesamt noch 26.495 Reviews berücksichtigt. Anhand des Sentiment Scores wurden drei Gruppen gebildet: positive (Sentiment Score ab 0,2), negative (Sentiment Score bis -0,2) und neutrale Reviews. Der Sentiment Score nimmt keine kategorische Einteilung vor und somit wurden für eine möglichst fehlerfreie Einteilung die positive und negative Gruppe nicht im Nullpunkt der Sentimentskala getrennt. Die Vorverarbeitung der Daten umfasste ebenfalls den Prozess der Tokenization, die Umwandlung in Kleinbuchstaben und das Herausfiltern von Stopwords. Zusätzlich wurde das „Term Frequency-Inverse Document Frequency"-Verfahren angewandt, um nicht untersuchungsrelevante Terme zu entfernen. Daraufhin wurden die Tokens auf ihre Stammform reduziert und in N-Gramme mit einer maximalen Länge von 3 zerlegt. Anschließend wurde ein Topic Modeling per LDA angewandt. Da kein hierarchisches LDA Tool zur Verfügung stand, wurden Verteilungen mit 5 bis 30 Themen getestet, um interpretierbare und möglichst aussagekräftige Topic Models zu erhalten. Dabei wurden in 1.000 Iterationen reproduzierbare Topic Models mit jeweils 15 zugehörigen assoziierten Termen (Top Words) pro Topic erzeugt. Für die Bestimmung der Hyperparameter α bzw. β wurde eine Heuristik angewandt und sie wurden automatisch iterativ optimiert, was der gängigen Praxis entspricht [28].

Um möglichen Qualitätsdefiziten des Datensatzes entgegenzuwirken, wurden weitere Ausschlusskriterien durchgesetzt. Als Ausgleich möglicher Ungenauigkeiten in der SA durch das domänenübergreifende Vorgehen wurden die Ratings als weiterer Indikator für die Einteilung der finalen Untersuchungsgruppen genutzt. Aus den positiven Reviews wurden danach nur noch Reviews mit 4 oder mehr Sternen, aus den negativen Reviews Einträge mit 2 oder weniger Sternen betrachtet.

4 Identifikation von Erfolgsfaktoren für AR-Applikationen

4.1 SA und Teilung des Datensatzes in Untersuchungsgruppen

Nach dem beschriebenen Verfahren wurde die Stichprobe anhand einer SA in drei Gruppen geteilt. Als Indikator für die Validität der durchgeführten SA wird zunächst der Zusammenhang zwischen Sentiment Score und dem vergebenen Rating des Reviews betrachtet. Das Rating wird auf einer Skala von 1 (schlechteste Bewertung) bis 5 (beste Bewertung) vergeben. Ein hoher Sentiment Score geht mit einem hohen Rating einher, r (26.495) = 0,377, p < 0,001 und Tabelle 1 zeigt klare deskriptive Unterschiede im durchschnittlichen Rating zwischen den Untersuchungsgruppen.

Tabelle 1. Rating und Sentiment Score der vorläufigen Untersuchungsgruppen

	Reviews in Stichprobe	*Negative Gruppe*	*Neutrale Gruppe*	*Positive Gruppe*
Anzahl	26.495	1.330	15.291	9.874
Durchschnittliches Rating	4,079	3,009	3,777	4,690
Durchschnittliches Sentiment	0,143	-0,357	0,05	0,354

Insgesamt weist dies auf eine erfolgreiche Einteilung des Sentiments hin. Allerdings enthält die negative Untersuchungsgruppe viele Reviews mit positivem Rating (siehe Abbildung 2). Da von einem positiven Inhalt eines Reviews mit 5 Sternen auszugehen ist, könnten diese einen störenden Einfluss auf die Ergebnisse nehmen. Daher wurde die positive bzw. negative Untersuchungsgruppe um Reviews mit zu niedrigen bzw. hohen Ratings gekürzt. Nach dieser Einteilung werden die Untersuchungsgruppen separat betrachtet und davon ausgegangen, dass ihr Inhalt Aufschluss über Erfolgsfaktoren von AR-Apps geben kann.

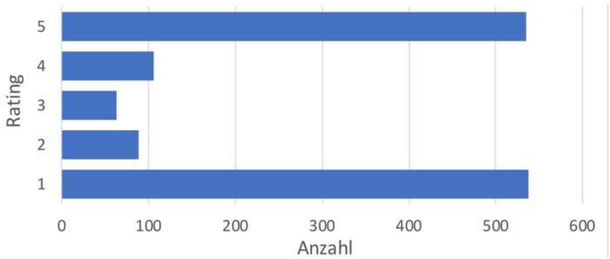


Abbildung 2. Rating der Reviews in negativer Untersuchungsgruppe

4.2 Topic Modeling in negativer Untersuchungsgruppe

Die Untersuchungsgruppe mit negativen Reviews umfasst nach dem oben beschriebenen Ausschlussverfahren 626 Datensätze, davon gehören 397 zur App „Pokémon Go“. Mit einem durchschnittlichen Rating von 2,669 und Sentiment Score von 0,014 im ungekürzten Datensatz gehört diese App nach beiden Kriterien zu den schlechtesten. Eine starke Repräsentation im negativen Datensatz ist damit nicht zu vermeiden, jedoch wird bei der Interpretation eine Abstrahierung von App-Inhalten angestrebt. Als Grundlage für die Auswertung dient ein per LDA generiertes Topic Model mit 10 vorgegebenen Topics (siehe Tabelle 3 im Anhang), wobei sinnvolle Top Words in Kombination mit hochassoziierten Reviews interpretiert werden.

Die Betrachtung dieses Topic Models fasst wiederkehrende Inhalte in negativen Reviews zusammen und macht somit Problemfelder deutlich. Hierunter befinden sich auch einige Themen, die keinen eindeutigen Bezug zu AR haben. So wird allgemeine Unzufriedenheit ausgedrückt mit beispielhaften Reviewinhalten wie „huge regret downloading“. Zudem wird das Preis-Leistungs-Verhältnis in negativen Reviews als nicht ausgeglichen wahrgenommen und die jeweilige App z.B. als „a waste of money“ bezeichnet. Der hohe Stellenwert von RM für Hersteller konkreter Anwendungen wird durch die direkte Adressierung der Entwickler und den Bezug auf bestimmte Updates und Softwareversionen deutlich. Hier werden fundamentale funktionale Probleme wie „cannot log in after update“ oder „new update makes this game unplayable“ zurückgemeldet. Dies bestätigt die besondere Rolle von Reviews, in denen Benutzer den Austausch mit den Verantwortlichen suchen, um Probleme anzumerken. Daher bietet RM vor allem für Entwicklungsteams einzelner Anwendungen die Möglichkeit, schnell auf Nutzerbeschwerden aufmerksam zu werden [22].

Zudem werden gehäuft Softwarefehler und Bugs, Server Probleme und App Crashs sowie ein hoher Batterieverbrauch bemängelt. AR-Apps analysieren die Umwelt durch ein Tracking System und benötigen daher Zugriff auf verschiedene Sensordaten. Die Erhebung und Verarbeitung dieser Daten in Echtzeit stellen hohe Anforderungen sowohl an die Hardware der mobilen Geräte, als auch an die Serverleistung [14]. Daher sollten diese Problemfelder als AR-bezogene Faktoren interpretiert werden.

Eine unmittelbare Relevanz für die Umsetzung von AR zeigt sich in der Beanstandung von fehlendem oder fehlerhaftem Tracking; beispielsweise wird „i think there is something wrong with the compass“ zurückgemeldet. Wie oben erläutert ist das Tracking der Umwelt zentrale Voraussetzung für die erfolgreiche Umsetzung eines AR Erlebnisses. Als fundamentales Problem wird zudem ein fehlender oder funktionsloser AR Modus bemängelt, häufig in Zusammenhang mit Kompatibilitätsproblemen bestimmter Geräte oder Betriebssystemversionen: „AR doesn’t work for the 7“.

Nach den Erkenntnissen dieses Topic Models liegen die Quellen von Unzufriedenheit bei Nutzern von AR-Apps vor allem in technischen Grundlagen.

4.3 Topic Modeling in positiver Untersuchungsgruppe

Nach der Filterung der Stichprobe anhand des Sentiment Scores und Ratings stehen noch 9.149 Reviews mit positiven Nutzereindrücken für die Untersuchung zur Verfügung. Diese Auswahl wird von Apps dominiert, die die Erkundung des Nacht- und Sternenhimmels thematisieren: Reviews zu Apps in diesem Themenbereich machen insgesamt 3.942 der Einträge in der Untersuchungsgruppe aus. Auch hier wird die Interpretation des Topic Models soweit wie möglich unabhängig von konkreten App-Inhalten durchgeführt. Die Darstellung der Ergebnisse basiert auf der Interpretation eines 20 Topics umfassenden Topic Models, das in Tabelle 4 im Anhang dargestellt ist. Darin werden erneut sinnvolle Top Words und assoziierte Reviewinhalte zur inhaltlichen Einordnung berücksichtigt.

Analog zur negativen Untersuchungsgruppe werden hier als AR-unabhängige Faktoren ein ausgewogenes Preis-Leistungs-Verhältnis gelobt und eine allgemeine Zufriedenheit und Empfehlung ausgesprochen: „Totally worth it!“, „I would advise this to all my friends“. Ebenso beziehen sich auch die positiven Rückmeldungen gehäuft auf bestimmte Updates und richten sich direkt an die Anwendungsentwickler, was sich in Reviewinhalten wie „The addition of the X-Ray vision improves this already brilliant app“ zeigt. Dabei werden sogar konkrete Verbesserungs- und Erweiterungsvorschläge gemacht, die in der Regel durch „Please add…“ oder „Can you add…“ eingeleitet werden. Dies macht erneut den hohen möglichen Nutzen von RM als Evaluationstool von spezifischen Apps deutlich.

Die Originalität und der Unterhaltungswert der Apps werden wiederholt positiv zurückgemeldet, z.B. mit „fun and addicting“ und „This is probably the most unique app in the App Store“. Während der reine Spaßfaktor keinen deutlichen AR-Bezug aufweist, kann die Neuartigkeit der Umsetzung von App-Inhalten mithilfe der AR-Technologie einen positiven Einfluss auf die wahrgenommene Originalität bedeuten. Besonders häufig wird zudem die erfolgreiche Umsetzung von Bildungsinhalten in positiven Nutzereindrücken festgehalten, z.B. „This app is making kids love astronomy“ und „all the information is very clear and to the point“. Dies könnte auf den hohen Anteil an Apps mit Bildungsinhalten zum Thema Himmel in der untersuchten Reviewgruppe zurückzuführen sein. Allerdings gehören edukative Anwendungen zu den besonders erfolgversprechenden Kategorien von AR-Apps, zu denen bereits weiterführende Studien zum Nutzerverhalten durchgeführt wurden [4]. Die besondere Eignung und fortgeschrittene Gestaltungsrichtlinien von AR für die Umsetzung von Lernumgebungen kann somit auch zur Erklärung der gehäuften positiven Reviews in diesem Bereich beitragen. Dies gibt Anlass dazu, experimentelle Nutzerdaten auch für andere Anwendungsbereiche zu erheben, um die Nutzererfahrungen auch hier anhand von evidenzbasierten, auf AR-Umgebungen zugeschnittenen Gestaltungsrichtlinien verbessern zu können.

Zentraler Inhalt positiver Rückmeldungen mit eindeutigem Bezug zu AR sind das Interface und die Benutzerfreundlichkeit, was die besonderen Möglichkeiten der AR-Technologie für die Implementierung intuitiver Schnittstellen unterstreicht. Reviews beziehen sich mit Aussagen wie „The way this all was designed and implemented seems incredible“ und „amazingly accurate and intuitive“ auf die grundsätzlichen Stärken der Interfaces. Darüberhinaus werden aber auch multimodale Ansätze gelobt („love this app and also the music“) und trotz spezifischem Bezug zur App „Pokémon Go“ wird der Wunsch nach lokaler Interaktion mit anderen Nutzern deutlich, was die Implementierung kollaborativer AR-Interfaces bekräftigt. Auch der AR-Effekt an sich wird von Nutzern positiv zurückgemeldet. Assoziierte Reviewinhalte wie „This app is my (…) window on the real sky“ und „makes pics come alive!“ zeugen von einer erfolgreichen Umsetzung der „magic lens“ nach Gervautz und Schmalstieg [36].
Zusammenfassend legt dieses Topic Model die Ausnutzung der vielfältigen UI-Gestaltungsmöglichkeiten als zentralen Adoptionstreiber dieser Technologie nahe.

5 Diskussion

Die Topic Models decken Quellen von Lob und Kritik der Nutzer von AR-Apps auf und lassen damit Rückschlüsse zu den Erfolgsfaktoren von AR-Apps zu. Das negative Topic Model verdeutlicht die wiederholte Beanstandung von technischen Problemen: die Umsetzung des Trackings, sowie Verfügbarkeits- und Kompatibilitätsprobleme. Hingegen werden Aspekte der Benutzerschnittstelle nicht kritisiert. Im Einklang damit heben die Ergebnisse des positiven Topic Models die Implementierung eines benutzerfreundlichen Interface als wesentlichen Adoptionstreiber hervor. Die

Ausschöpfung der Interaktions- und Darstellungsmöglichkeiten der AR-Technologie kann somit eine vorteilhafte Mensch-Computer-Schnittstelle erreichen.

Die Durchführung der vorgestellten Untersuchung ist mit Limitationen verbunden. Das Format von App Store Reviews geht mit einer geringen inhaltlichen Qualität der Datengrundlage einher, die in der Verarbeitung und Analyse kompensiert werden muss, sodass trotzdem aufschlussreiche Erkenntnisse gewonnen werden können [22], [37]. Im Hinblick auf die inhaltliche Datenqualität könnten die Untersuchung ergänzt werden, indem die bewertete Nützlichkeit der Reviews berücksichtigt wird. So könnten Reviews, die von mindestens fünf anderen Nutzern als „hilfreich" bewertet wurden, gesondert ausgewertet werden. Darüber hinaus könnten abwägende Reviews mit mittleren Bewertungen und ohne Zuordnung zu den polaren Untersuchungsgruppen konstruktive Kritik enthalten und sollten daher ebenfalls betrachtet werden.

Wie oben bereits erwähnt, bestätigt die Untersuchung einen hohen Nutzen von RM einzelner Apps für Anwendungsentwickler. Damit können Trends in den Nutzerrückmeldungen schnell erkannt und entsprechende Anpassungen umgesetzt werden. Im Rahmen dieser Untersuchung wurde jedoch angestrebt, allgemeine Schlussfolgerungen in Bezug auf bestimmte App-Eigenschaften zu ziehen. Um dies erfolgreich umzusetzen, hat ein ausgeglichener Datensatz einen hohen Stellenwert. Im hier untersuchten Fall der AR-Apps stehen zum Zeitpunkt der Untersuchung jedoch noch nicht genug Reviews in vielfältigen AR-App-Kategorien zur Verfügung, sodass einzelne Anwendungen wie „Pokémon Go" und „Sky Guide" die Analyseergebnisse dominieren. Sobald eine ausreichende Datengrundlage besteht, könnten weiterführende Untersuchungen eine Aufgliederung der Analyse nach den oben angegebenen Kategorien von AR-Apps vornehmen. Darüber hinaus adressieren die identifizierten Faktoren nicht nur spezifische Eigenschaften von AR, sondern auch allgemeine Aspekte, die es in der App-Entwicklung zu berücksichtigen gilt.

Auf methodischer Ebene würde die Interpretierbarkeit der Ergebnisse von einer SA auf dem Level von Aspekten profitieren, sodass verschiedene inhaltliche Bestandteile eines Reviews einzeln klassifiziert werden. Zudem würde die Anlage und Verwendung eines domänenspezifischen Training Sets die Treffsicherheit der Einteilung erhöhen. Das Topic Modeling wurde in dieser Untersuchung mittels eines klassischen LDA Verfahrens durchgeführt, da in der gewählten Data Analyse-Umgebung kein kostenfreies hierarchisches LDA Verfahren verfügbar war. Damit könnten allerdings die Trennschärfe zwischen den Topics und die allgemeine Interpretierbarkeit des Topic Models erhöht werden. Aufgrund der grundsätzlich geltenden eingeschränkten inhaltlichen Qualität von App Reviews und methodischer Einschränkungen kann ein RM in keinem Fall die Ergebnisse einer wissenschaftlichen Nutzerstudie ersetzen.

6 Fazit und Ausblick

Durch die Durchführung eines RM wurden Rezensionen über AR-Apps analysiert. Die Untersuchung trägt mit Erkenntnissen zu den Erfolgsfaktoren von AR-Apps zur Erweiterung der Wissensbasis im Bereich der Applikationsentwicklung bei. Durch die Auswertung von Nutzerbewertungen begegnet diese Untersuchung einer aktuellen

Forschungslücke in der Betrachtung der User Experience von AR-Technologien [38]. Mithilfe von SA wurden Topic Models zu positiven und negativen Benutzerrückmeldungen generiert. Dabei werden vor allem Defizite in den technischen Grundlagen wie dem Tracking System und der Gerätekompatibilität als Entwicklungsfeld identifiziert. Eindeutiger Adoptionstreiber von AR-Apps ist die Umsetzung eines intuitiven Interface mithilfe der AR-Technologie. Die erfolgreiche Anwendung dieses Vorgehens schließt an den von Panichella et al [39] skizzierten Forschungsbedarf zum Einsatz von Topic Modeling in der Reviewauswertung an.

Die Analyse und Auswertung wurde von Wissenschaftlern aus Sicht der App-Entwicklung vorgenommen. Die vorausgehende Erfahrung in der Entwicklung eigener Applikationen stärkt die Validität der Ergebnisse (vgl. [37]). Die identifizierten Erfolgsfaktoren tragen zur Wissensbasis der Entwicklung von AR-Apps bei und sind damit von wissenschaftlicher und praktischer Relevanz. Die gewonnenen Erkenntnisse tragen Implikationen zur Erforschung des Nutzererlebnisses von AR-Applikationen bei (vgl. [40]). Weiterhin unterstützen die vorgestellten Erkenntnisse die erfolgreiche Optimierung und Neuentwicklung von AR-Applikationen.

Die identifizierten Faktoren adressieren auch Gestaltungsempfehlungen, die nicht AR-spezifisch sind. Dies impliziert die weiterhin gegebene Relevanz grundsätzlicher Charakteristika der App-Entwicklung für AR-Anwendungen. Im Einklang mit Forderungen der Wissenschaft sollte darauf aufbauend eine Erweiterung und Adaption vorhandener Gestaltungsrichtlinien für spezifische Technologien vorgenommen werden (vgl. [41]). Die spezifischen AR-Faktoren liefern dazu wertvolle Implikationen für die Entwicklung einer Gestaltungstheorie für AR-Applikationen. Die Ergebnisse können unter anderem im Rahmen von Design Science Research im Bereich der Entwicklung von AR-Anwendungen genutzt und ausgebaut werden [42], [43]. Die Erkenntnisse können durch eine Erweiterung der Analyse auf Nutzerreviews von weiteren Plattformen und neuerscheinenden Applikationen ausgebaut werden. Weiterer Forschungsbedarf besteht hinsichtlich der Übertragbarkeit der Erkenntnisse auf AR-Apps für andere Endgeräte, beispielsweise AR-Glasses.

Literatur

1. Apple Inc.: Introducing ARKit: Augmented Reality for iOS.
2. Google LLC: Introducing ARCore: Augmented Reality at Android Scale (GDD Europe '17).
3. Sutherland, I.E.: The ultimate display. In: Proceedings of the Congress of the Internation Federation of Information Processing (IFIP). pp. 506–508 (1965).
4. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review. Twenty-third Am. Conf. Inf. Syst. 1–10 (2017).
5. Irshad, S., Rambli, D.R.A.: Advances in Mobile Augmented Reality from User Experience Perspective: A Review of Studies. In: Advances in Visual Informatics: 5th International Visual Informatics Conference, IVIC 2017, Bangi, Malaysia, November 28--30, 2017, Proceedings. pp. 466–477 (2017).
6. Kourouthanassis, P.E., Boletsis, C., Lekakos, G.: Demystifying the design of mobile augmented reality applications. Multimed. Tools Appl. 74, 1045–1066 (2015).
7. Dey, A., Billinghurst, M., Lindeman, R.W., Swan, J.E.: A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. Front. Robot. AI. 5, 1–28 (2018).

8. Shitkova, M., Holler, J., Heide, T., Clever, N., Becker, J.: Towards Usability Guidelines for Mobile Websites and Applications. Wi. 1603–1617 (2015).
9. Azuma, R.T.A.: Survey of Augmented Reality, (1997).
10. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. 21, 34–47 (2001).
11. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. Telemanipulator Telepresence Technol. 2351, 282–292 (1994).
12. Olsson, T., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: A user study in the context of shopping centres. Pers. Ubiquitous Comput. 17, 287–304 (2013).
13. Kim, K., Hwang, J., Zo, H., Lee, H.: Understanding users' continuance intention toward smartphone augmented reality applications. Inf. Dev. 1–14 (2014).
14. Mekni, M., Lemieux, A.: Augmented Reality : Applications , Challenges and Future Trends. Appl. Comput. Sci. anywhere,. 205–214 (2014).
15. Manuri, F., Sanna, A.: A Survey on Applications of Augmented Reality. Adv. Comput. Sci. 5, 18–27 (2016).
16. Billinghurst, M., Clark, A., Lee, G.: A Survey of Augmented Reality. Found. Trends® Human–Computer Interact. 8, 73–272 (2015).
17. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., Ivkovic, M.: Augmented reality technologies, systems and applications. Multimed. Tools Appl. 51, 341–377 (2011).
18. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits, and atoms. Proc. ACM SIGCHI Conf. Hum. factors Comput. Syst. 234–241 (1997).
19. Zhou, F., Dun, H.B.L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In: Proceedings - 7th IEEE International Symposium on Mixed and Augmented Reality 2008, ISMAR 2008. pp. 193–202 (2008).
20. Reitmayr, G., Schmalstieg, D.: Mobile collaborative augmented reality. In: Proceedings - IEEE and ACM International Symposium on Augmented Reality, ISAR 2001. pp. 114–123 (2001).
21. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Found. Trends® Inf. Retr. Vol. 2, No 1-2 1-135. 2, 1–135 (2008).
22. Genc-Nayebi, N., Abran, A.: A systematic literature review: Opinion mining studies from mobile app store user reviews. J. Syst. Softw. 125, 207–219 (2017).
23. Martin, W., Sarro, F., Jia, Y., Zhang, Y., Harman, M.: A Survey of App Store Analysis for Software Engineering. 43, 817–847 (2017).
24. Khan, K., Baharudin, B., Khan, A., Ullah, A.: Mining opinion components from unstructured reviews: A review. J. King Saud Univ. - Comput. Inf. Sci. 26, 258–275 (2014).
25. Tsytsarau, M., Palpanas, T.: Survey on mining subjective data on the web. Data Min. Knowl. Discov. 24, 478–514 (2012).
26. Blitzer, J., Dredze, M., Pereira, F.: Domain adaptation for sentiment classification. Annu. Meet. Comput. Linguist. 45, 440–447 (2007).
27. Yadollahi, A., Shahraki, A.G., Zaiane, O.R.: Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Comput. Surv. 50, 1–33 (2017).
28. Debortoli, S., Müller, O., Junglas, I., Vom Brocke, J.: Text Mining For Information Systems Researchers: An Annotated Topic Modeling Tutorial. Commun. Assoc. Inf. Syst. 39, (2016).
29. Rossetti, M., Stella, F., Zanker, M.: Analyzing user reviews in tourism with topic models. Inf. Technol. Tour. (2016).
30. Blei, D., Ng, A.: Latent dirichlet allocation. J. Mach. Learn. 3, 993–1022 (2003).
31. Blei, D.M.: Probabilistic topic models. Commun. ACM. 55, 77–84 (2012).

32. Blei, D., Griffiths, T.L., Jordan, M.I., Tenenbaum, J.B.: Hierarchical topic models and the nested Chinese restaurant process. Adv. Neural Inf. Process. Syst. 16, 106 (2004).
33. Crystal, D.: A global language. In: English in the World. pp. 151–196 (2013).
34. Bye, K.: Scrape iOS app review data from iTunes.
35. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC. 10, 2200–2204 (2010).
36. Gervautz, M., Schmalstieg, D.: Anywhere interfaces using handheld augmented reality. Computer (Long. Beach. Calif). 45, 26–31 (2012).
37. Chen, N., Lin, J., Hoi, S.C.H., Chen, N., Lin, J., Hoi, S.C.H., Xiao, X., Zhang, B.: AR-Miner : Mining Informative Reviews for Developers from Mobile App Marketplace AR-Miner : Mining Informative Reviews for Developers from Mobile App Marketplace. (2014).
38. Parvinen, P., Hart, A., Griswold, M., Lai, J., Boyer, E.W., Skolnik, A.B., Chai, P.R.: Introduction to Minitrack : Mixed , Augmented and Virtual Reality. 1395–1396 (2018).
39. Panichella, S., Sorbo, A. Di, Guzman, E., Visaggio, C.A., Canfora, G., Gall, H.C.: How Can I Improve My App ? Classifying User Reviews for Software Maintenance and Evolution. 2015 IEEE Int. Conf. Softw. Maint. Evol. 281–290 (2015).
40. Niemöller, C., Metzger, D., Thomas, O.: Design and Evaluation of a Smart-Glasses-based Service Support System. In: 13. International Conference on Wirtschaftsinformatik (WI). pp. 106–120 (2017).
41. Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N.: Designing the User Interface: Strategies for Effective Human-Computer Interaction. Pearson (2016).
42. Metzger, D., Niemöller, C., Jannaber, S., Berkemeier, L., Brenning, L., Thomas, O.: The Next Generation: Design and Implementation of a Smart Glasses-based Modelling System. Enterp. Model. Inf. Syst. Archit. - Int. J. Concept. Model. [accepted for publication] (2018).
43. Zobel, B., Berkemeier, L., Werning, S., Vogel, J., Ickerott, I., Thomas, O.: Towards a modular reference architecture for smart glasses-based systems in the logistics domain. In: CEUR Workshop Proceedings. p. Paper 17 (2018).

Anhang

Tabelle 2. Deskriptive Daten zu gescrapten Reviews und Untersuchungsgruppen

Untersuchte App	*Bewertung*	*Reviews*	*Reviews in negativer Gruppe*	*Reviews in positiver Gruppe*
Sky Guide	4,9	10.884	23	2.865
Pokémon Go	3,7	86.217	397	433
Egg, Inc.	4,8	21.666	14	1.219
Zombie Gunship Revenant AR	4,7	415	6	56
Star Walk 2 – Night Sky Map	4,8	2.559	17	1.200
CamToPlan PRO	4,6	527	8	215
Tape Measure	4,2	153	9	19
Star Walk 2 Ads+: Sky Guide	4,8	408	10	142

Untersuchte App	*Bewertung*	*Reviews*	*Reviews in negativer Gruppe*	*Reviews in positiver Gruppe*
Lumyer	4,7	2.898	14	1.710
magicplan	4,7	2.045	38	672
Thomas & Friends Minis	4,5	168	3	32
AirMeasure – AR Tape & Ruler	4,6	95	3	20
Euclidean Lands	4,7	155	7	19
Human Anatomy Atlas 2018	4,9	163	0	56
Stack AR	3,8	472	47	36
Monster Park – AR Dino World	4,5	90	6	21
Heads Up Navigator	2,3	37	3	6
Plotaverse Photo Video Editor	4,5	471	6	90
Inkhunter Try Tattoo Designs	4,7	782	12	329
GIPHY World: AR GIF Stickers	4,6	57	3	9

Tabelle 3. Topic Model zu negativer Untersuchungsgruppe mit 10 Topics

Themen-bereich	*Inhaltliche Interpretation*	*Top Words*	*Assoziierte Reviewinhalte*
Preis-Leistungs-Verhältnis, Empfehlung	Allgemeine Unzufriedenheit	download, delete, useless	„huge regret downloading“, „won’t download“, „The game is useless…“
	Preis-Leistungs-Verhältnis	waste_time, money, waste	„It’s a waste of money“, „This is so bad I do not like this I want my money back“
Direkte Kommu-nikation	Direkte Entwickler-kommunikation	company, customer, user	„A shameless cash grab from a company with a poor track record from listening to their customers.“
	Probleme nach Update	worst_update, ruin, unplayable	„New updated makes this game unplayable, not playing until it is fixed“
	Login Probleme nach Update	update, broken, account	„Cannot log in after update“

Themen-bereich	*Inhaltliche Interpretation*	*Top Words*	*Assoziierte Reviewinhalte*
Technische Probleme	Softwarefehler, Bug	play, freeze, problem	„lag and freeze anytime“, „freezes, hard to use“
	Server Probleme, App Crashs	crash, server, keep_crash	„This game is crashing since i got it…“, „Needs More Servers“
	Batterie-verbrauch	bad, battery	„has now become a huge drain on my battery life since the battery saver removal“
	Tracking System, Sensoren	compass, wrong	„I think there is something wrong with the compass in all iphone 6s models“,
	AR Modus, Kompatibilität	won't, work, ar_mode, ar_work	„Very stupid how the AR doesn't work for the 7“, „AR doesn't work, no visual feedback in order to track the surface“

Tabelle 4. Topic Model zu positiver Untersuchungsgruppe mit 20 Topics

Themen-bereich	*Inhaltliche Interpretation*	*Top Words*	*Assoziierte Reviewinhalte*
Preis-Leistungs-Verhältnis, Empfehlung	Preis-Leistungs-Verhältnis	worth, purchase, price	„Well worth 1.49… probably more like $5.00“, „Totally worth it!“
	Preis-Leistungs-Verhältnis, Empfehlung	definitely_worth, thumbs	„Great app, definitely worth the 2.99“, „Definitely recommend this app“
	Allgemeine Zufriedenheit, Empfehlung	amazing, favorite, friends, familiy	„What an amazing app this is… I have shared it with all my grandkids.“, „I would advise this to all my friends“
Direkte Kommu-nikation	Verbesserung nach Update	version, feature, update, improve	„The addition of the X-Ray vision improves this already brilliant app“
	Direkte Entwickler-kommunikation	work, keep, update	„Love it. Now they need to keep building on it!“, „I can't wait for the next update! When's it coming out?“
	Direkte Entwickler-kommunikation	please, add, option	„Please add…“, „Can you add…“

Themenbereich	*Inhaltliche Interpretation*	*Top Words*	*Assoziierte Reviewinhalte*
Originalität, Unterhaltung	Spaß	addictive, fun_play	„Fun and addicting“, „I haven’t stopped playing for days“
	Originalität	nice, clever, unique	„This is probably the coolest and most unique app in the App Store“
	Originalität und Spaß	time, idea	„play it all the time“, „Great idea“, „the app is definitely ONE OF A KIND!“
Bildungsinhalte	Bildungsinhalte	learn, help, know	„An outstanding app! This really helps me find and learn more about the stars…“
	Bildungsinhalte	learn, tool, kids	„I got it for my daughter because she wants to know more about the universe“, „This app is making kids love astronomy“
	Bildungsinhalte	information, detailed, accurate	„All the information is very clear and to the point“, „Chock full of information at one’s fingertips“
	Bildungsinhalte, Interface	beautiful, graphics, informative, educational	„A gorgeous, fluid interface“, „This app is so beautiful and educational!“
Benutzerfreundlichkeit	Benutzerfreundlichkeit, Effektivität	user_friendly, easier	„can make your work that much easier“, „awesome, user friendly“
	Interface	impressive, design, quality	„The way this all was designed and implemented seems incredible.“
	Interface	easy, fun, interface	„Great job on the UI, really easy to use“, „amazingly accurate and intuitive“
	Kollaboratives Interface	trade, gym, team	„I would love to see the day when you can chose a battle with another local user and also being able to trade your Pokémon“
	Mulitmodales Interface	relaxing, music	„love this app and also music – so relaxing“
Immersion	Immersion	look, see, show	„This app is my only window on the real sky“, „whats better than to look up and map what you see with what’s on your phone. Magical“
	Immersion	effects, life, creative	„fun photoapp that makes a picture more alive“, „makes pics come alive! So creative!“

Designing Dynamic Decision Support for Electronic Requirements Negotiations

Annika Lenz[1]

[1] University of Hohenheim, Department of Information Systems I, Stuttgart, Germany
annika.lenz@uni-hohenheim.de

Abstract. Decision support in software development is particularly important for requirements negotiations to help assessing requirements and their different implementation alternatives. Changes related to requirements are likely, which impede decision making. To keep an overview of the assessment of requirements and to keep this assessment up-to-date throughout a software development project, flexible decision support processes are needed. In this paper, we design interactive dynamic decision support, which can handle changes related to requirements dynamically. The designed support component is compared to two state-of-the-art approaches for decision support in requirements negotiations.

Keywords: Dynamic decision support, requirements negotiation, dynamic preference measurement

1 Introduction

To keep an overview of the various requirements in large software development projects is a great challenge, even if requirements management systems are employed. It gets even more challenging, if negotiations must be conducted, to determine, which requirements are actually implemented and how. Requirements negotiation is an "iterative process of communication and decision-making between [stakeholders] who have the overall goal of agreeing on a software development process and outcome" [1, p. 304]. In such coordination and reconciliation processes, requirements to be implemented, their development cost, and the delivery schedule are negotiated [2].

Decision support in such scenarios enables the quantification of requirements and possible alternatives of their implementations and thus makes them comparable [3]. In negotiations, in which various stakeholders (e.g. customers, developers, project managers, product owners) have a say, knowing one's own position and expressing this position quantitatively provides enormous benefits to enforce one's own interests.

A drawback is that software development is an industry characterised by a high degree of dynamics [4]. Especially in large software development projects, new information regarding requirements is obtained or disclosed throughout the project, so changes in requirements are likely [5]. An existing set of requirements must be refined in subsequent iterations of the requirements negotiation process in terms of additional requirements, omitted requirements, and/or changes in existing requirements' scope [2,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

6], which illustrates the high degree of uncertainty and incompleteness of information [7].

Moreover, negotiations underlie process dynamics which influence the negotiators' preferences [8, 9]. In general, preferences are unstable due to new preference formation, learning or fatigue [10]. This effect is amplified by the negotiation partners' exchange during the negotiation process. By the negotiation partners' reciprocal behaviour they gain new information or achieve clarity about the negotiation issues or values, which influences their preferences [11].

Thus, preference changes in requirements negotiations may stem from agenda changes [12] or from the reciprocity of the negotiation process itself. In both cases, existing measurement of preferences must be efficiently adjusted to provide meaningful decision support. Dynamic decision support takes the perspective of time into consideration. It considers an unstable negotiation agenda (i.e. requirements or their implementation solution changes), and unstable preferences themselves.

From a requirements engineering perspective, decision making in requirements negotiations is supported by multi-criteria decision analysis approaches [e.g. 13], which methodologically consider a dynamic perspective in their requirements negotiation process, however, do not carry it to the decision support process. Negotiation research delivers decision support dealing with incomplete information [e.g. 14]. However, literature, which integrates insights of both domains is still scarce [2, 15]. There is no comprehensive approach that focuses on supporting decisions in requirements negotiations on a dynamic, interactive level to allow for adjustment of preferences [cf. 15], although the necessity of dynamic methodologies for requirements negotiations is present [e.g. 6, 11].

Against this background, our aim is to design efficient dynamic decision support in requirement negotiations integrating a requirements engineering and negotiation research perspective. Dynamic decision support enables continuously accurate preferences, which allows an accurate negotiation analysis [16], i.e. analysis of own and the negotiation partner's requirements specification offers and concessions. To this end, the aim of our paper is two-fold: (1) to design a dynamic decision support component for requirements negotiations; (2) to evaluate its suitability by comparing it with state-of-the-art decision support in electronic requirements negotiations using a scenario-based approach [17, 18].

2 State-of-the-Art

In negotiations, the two perspectives of individual decision making and plural decision making are relevant [16]. Decision analysis takes an individual perspective and focuses on a normative and prescriptive approach for an individual negotiator or a single negotiation party, while negotiation theory as a perspective of plural decision making assumes joint decision making. Negotiation theory describes how real people could make better collaborative decisions. Both perspectives are integrated in asymmetric negotiation analysis, which supports a negotiator's position considering the partner's behaviour [14] to support achieving the best outcome [16]. Thus, decision support takes an individual decision making perspective in a joint decision making context, whose result and benefit is asymmetric negotiation analysis.

Prerequisite is that the individual preferences of the negotiating parties are measured beforehand. A multitude of preference measurement methods exists to quantify the decision makers' preferences. As terminology differs in the field of preference measurement methods, we refer to measuring preferences for attributes and alternatives. In self-explicated approaches [19], the decision makers elicit their preferences directly. The preference measurement can thus be carried out quickly and easily and the cognitive complexity can be kept low for the decision maker. For example, Adaptive Self-Explication (ASE, [20]) comprises of three different user dialogue steps: (1) the alternatives are rated on a scale from 0 to 1; (2) the attributes are ranked regarding their importance; (3) pairwise comparisons of attributes are conducted, for which two attributes are presented to the decision maker, who enters a ratio of how much more important one attribute is compared to the other. Another popular preference measurement methods is TOPSIS [21], which uses linguistic terms to determine the attribute weightings.

Electronic negotiation systems aim to support the three phases of a negotiation, namely the planning phase (or pre-negotiation or preparation), the actual negotiation phase, and the post-settlement phase. Negotiation support systems have an inherent characteristic of providing decision support [22]. They help decision makers to understand the problem and assess the implications of their decisions. To do so, preference measurement methods are applied to calculate the utility of single offers using multi-attribute utility theory [23] based on the importance weightings of negotiation issues (= attributes) and issue values (= alternatives) [e.g. 24–26]. The resulting utility model forms the basis for negotiation analysis [16].

Requirements engineering distinguishes between requirements and their implementation solutions [17]. Using the terminology of preference measurement, requirements refer to attributes and solutions refer to alternatives. In requirements negotiations, three types of issues are negotiated: i) design related issues, ii) contract related issues, iii) technological issues and for each implementation solutions respectively options to resolve the issue [2, 27, 28]. For the sake of better reading, we refer to all types as requirements and their solutions.

Involved negotiation parties cover success critical stakeholders of the software development project [6], who typically pursue mismatching goals: Future end users desire numerous features, a high service level, or fast delivery, buyers may also wish timely delivery, however are also interested in cost within budget, or compliance, while developers prefer stable requirements and flexible contracts [2]. Typically, only a relevant subset of these stakeholders are actually involved in the respective negotiation, which results in different negotiation constellations [29]. Although requirements negotiation is an iterative process [2], which is not a one-time episode but an on-going process in general, its instantiation characteristics depend on several factors such as the software development method deployed (e.g. traditional, agile, or hybrid software development), the project type, the collaboration situation [2], involved stakeholders [29], the project stage (beginning or completion of the project), negotiation scope (i.e. requirements and solutions, technologies, or contract aspects), or frequency. Thus, their structure may differ. Requirements negotiations can be performed in independent phases, each including respective stakeholders and resulting in a (partial) agreement [29].

Nonetheless, in almost every software development project, requirements negotiations are present in some way, and in each requirements negotiation, decisions must be made. Decision support accordingly is required for stakeholders' conflicting requirements perspectives in the actual negotiation phase. We consider negotiations, in which preference measurement is supported electronically and analysis is provided based thereon. Preference measurement can jointly be conducted in face-to-face negotiation workshops [e.g. 6] or be applied asynchronously for an individual decision maker to prepare the requirements negotiation [e.g. 30].

The most widely used methodology in the area of interactive requirements negotiation support is the WinWin methodology [6, 15], which aims to achieve a fair agreement among all involved stakeholders by attempting to meet the win conditions of each stakeholder [6]. Dealing with various stakeholders, group recommendation technologies, such as IntelliReq [31], dedicatedly aim to support collaboration and enable group decision making [15]. Based on the stakeholders' preferences, recommendation technologies are applied to reach a joint decision on which requirements to implement (first).

The presented existing approaches follow a dynamic paradigm and provide decision support, however dedicated preference adjustment mechanisms are missing. Applying dynamic preference measurement to the domain of requirements negotiations can improve such approaches substantially. Our work addresses the depicted gap by designing and evaluating such a concept.

3 Methodology

The methodology applied in this paper covers the design of dynamic decision support for requirements negotiations, DynaDeS, and the evaluation of the designed component, both based on scenarios to be captured [17, 18, 32].

We design the component to fulfil the goal of providing dynamic decision support to handle preference changes efficiently. To do so, scenarios are considered, for which the design goal must be fulfilled [17, 18]. The scenarios affecting preference information changes, which must be supported, stem from volatile requirements [2] and negotiation process dynamics [8, 11], see Table 1. When requirements or solutions change during the negotiation process, the decision support component is intended to provide a process to adjust their preferences accordingly. Thus, it must consider (A) a new requirement with new solutions, (B) a new solution without a new requirement and adjust their preferences. Moreover, preferences are unstable and may change over time despite of requirements or solution changes, thus the decision support component must handle (C) preference changes for requirements and (D) preference changes for solutions.

To reach the overall aim of handling preference changes efficiently, the design process conducts the following steps: (1) Existing decision support of negotiation research is adapted to requirements negotiations; (2) the preference measurement is extended to allow dynamic preference adjustment; (3) the decision support processes are adapted to meaningfully use the adjusted preferences.

Table 1. Scope of preference adjustment for dynamic decision support in requirements negotiations

#	*Scenario*	*Cause*
A	New requirement	Unstable requirements
B	New solution	Unstable requirements
C	Changed requirement preference	Negotiation process dynamics
D	Changed solution preference	Negotiation process dynamics

The designed DynaDeS is implemented in the web-based negotiation support system Negoisst [26, 30] to exploit the benefits of an integrated system. DynaDeS is evaluated by means of a scenario-based [17, 32] comparison to analyse whether the designed dynamic decision support component meets its design goals for requirements negotiations. Scenarios illustrate how design goals are satisfied [17]. Thus, we compare DynaDeS to existing decision support w.r.t. to their suitability for the required scenarios (cf. Table 1). Furthermore, as the WinWin methodology is the methodology most commonly used for electronic requirements negotiations [15], we include decision support within the WinWin methodology [6, 33]. Moreover, we compare IntelliReq [31], which provides dedicated group decision support in requirements negotiations. Although the support components to compare are heterogeneous in different characteristics, they coincide in taking a dynamic perspective on decision support for requirements negotiations.

4 Design of the Dynamic Decision Support Component

The decision support component is not intended to replace requirements elicitation processes or requirements refinement processes. It starts when an initial set of requirements and their implementation solutions is determined. This set may be rudimentary at the beginning and be refined during the superordinate negotiation process, however, requirements and solution refinement is not meant to be part of the decision support component. In contrast, the decision support component is supposed to handle the required scenarios of changes of the environment.

4.1 A Concept for Dynamic Decision Support

In the following, we will (1) adapt existing decision support of negotiation research to requirements negotiations; (2) enable dynamic preference adjustment; (3) design processes to meaningfully use the adjusted preferences.

(1) For decision support during requirements negotiations without a dynamic perspective, information about the agenda (i.e. requirements and solutions to negotiate) is gathered in the planning phase. The problem definition takes place based on the

project goals and the project phase [cf. 34]. At a very early stage, requirements negotiation involves negotiation issues on a high-level, while negotiation in a later stage of the software development project focuses on specific aspects or sub-projects. Requirements that are elicited at an early stage allow for a wide range of implementation solutions that will be specified later. Furthermore, in the planning phase success critical stakeholders' individual goals are collected. These goals may concern high-level issues, general system functions, budget, schedule, or technical issues, such as the development environment. Having extracted the initial agenda to negotiate, the negotiation issues and values are elicited. Subsequently, still during the planning phase, an initial preference elicitation is performed, resulting in a (tentative) utility model, which is required to provide asymmetric analysis in the actual negotiation phase.

During the actual negotiation phase, the agreement takes place. The stakeholders exchange offers and counteroffers to find a mutually beneficial and acceptable solution. Analytical support is provided to assess own offers or offer drafts as well as the negotiation partner's offers based on the preferences measured.

(2) To provide dynamic preference adjustment, we choose ASE [20] as basis preference measurement method, since it performs very well regarding its efficiency, validity, and cognitive complexity [35]. Furthermore, although it is a one-shot preference measurement method as is common in multi-criteria decision analysis, it provides the potential to be adopted for dynamic contexts [35], since it uses scarcely dependencies between preference information and allows to separate preference information measured and a utility model calculated based thereon. In this paper, we incorporate DynASE, which extends ASE by dynamic preference mechanisms, into DynaDeS.

For the dynamic preference adjustment, similar user dialogues as in the initial preference elicitation are applied, namely rating of the desirability of all solutions, ranking of the requirements according to their importance, and pairwise comparison of all requirements [20]. After a bundle of pairwise comparisons is conducted, a utility model can be calculated. An adaptive selection of pairwise comparisons facilitates to ask for the greatest possible information gain. If the resulting utility model is sufficiently precise, the preference adjustment process terminates. Otherwise, further pairwise comparisons are conducted until a pre-defined number of comparisons is reached, or all issues are compared at least once.

(3) The decision support process is adapted to react to negotiation process dynamics and to embed subsequent activities in the actual negotiation phase. It allows to adjust preferences and to re-calculate the utility model during the actual negotiation phase and shifts thereby activities of the planning phase into the negotiation phase (similar to agenda negotiations, cf. [12]), see Figure 1. Focus of the dynamic preference adjustment is the interaction of preference measurement and the utility model calculation. Most preference measurement methods do not separate the pure measurement of preferences from the calculation of the utility model. However, to enable efficient dynamic decision support, the possibility of separation is inevitable. The preferences resulting from the initial preference elicitation must be saved

independently of the utility model calculated to reuse still valid preferences in the case of changes.

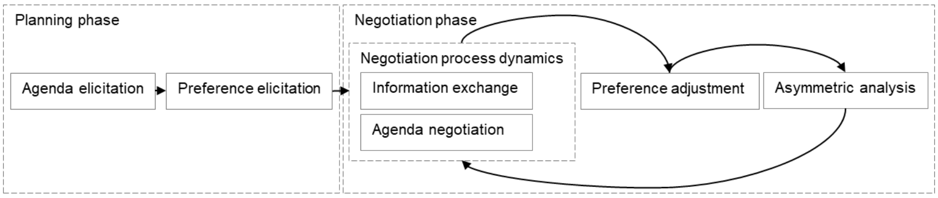


Figure 1. Dynamic decision support in the negotiation phases

In DynaDeS, the dynamic aspect comes into effect if new information is obtained, which result in (A) requirements, (B) solutions, (C) requirement preference, and/or (D) solution preference changes, see Figure 2. The planning phase is entered again in all scenarios. In the scenarios of A and B, additionally the agenda is adjusted. Subsequently in all scenarios, preferences are adjusted as well, dependent on the information gained: Solution ratings are maintained from the existing preference information if it is still valid. This can be facilitated because the solution ratings of one requirement do not relate to the ratings of other solutions. Except for scenario C when only requirement preference information changes, the first step of the preference measurement is repeated for (A) solutions of the new requirement, (B) a new solution, or (D) the outdated solution preference information. In the scenarios B and D where no requirement information has changed, the user dialogue is already completed and a renewed accurate utility model is calculated based on existing valid preference information and newly gathered preference information, which can be used in the stakeholder's analysis. For the scenarios of A and C, the preference adjustment process continues, presenting the existing requirements ranking to the user, in which (A) the new requirement must be sorted respectively (C) the rank of the respective requirement must be corrected. Preference information of paired comparisons of all other requirements is reused. Thus, based on the ranking and valid paired comparisons, a renewed utility model is calculated again. If the utility model is not yet sufficiently precise, additional pairwise comparisons involving the respective requirement can be performed. Optionally, an interactive check allows the stakeholders to view and validate their utility model and initiate preference adjustment in case the utility model does not fit their preferences (anymore). Again, based on a newly calculated utility model, precise analytical support can be provided.

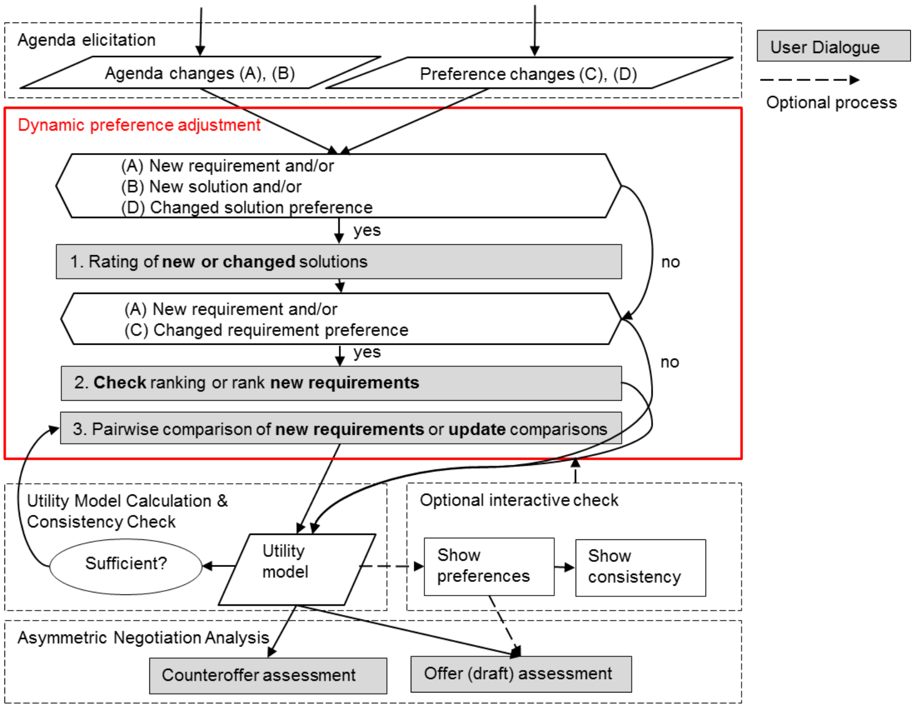


Figure 2. DynaDeS

4.2 Instantiation of Dynamic Decision Support in a Negotiation Support System

The designed dynamic decision support is implemented in the negotiation support system Negoisst [26, 30] as it already offers holistic negotiation support. Negoisst is a web-based support system, which provides decision support, communication support, and document management. One aspect of its communication-orientation is to support the offer exchange by semi-structured message exchange, which allows to send unstructured text messages along with a structured, formal offer. To make binding offers, the negotiators send each other semi-structured messages in an alternating order.

Negoisst provides extensive analytical features for decision support [e.g. 14]. The analytical features enable to assess offers quantitatively based on the negotiators' preferences by utility values. The utility of offers can be analysed by each negotiator. The history graph serves as visual support of the negotiators' convergence during the negotiation process, depicting the utility of offers over time. Concessions and gains can be analysed visually. Negoisst provides analytical support during the phases of planning (e.g. preference elicitation) and in the phase of negotiation (e.g. offer assessment and analysis). The implementation of the preference adjustment within DynaDeS covers the following three user dialogues: (1) rating of solutions; (2) ranking of requirements; (3) comparing requirements pairwise by using a slider or input fields, see Figure 3.

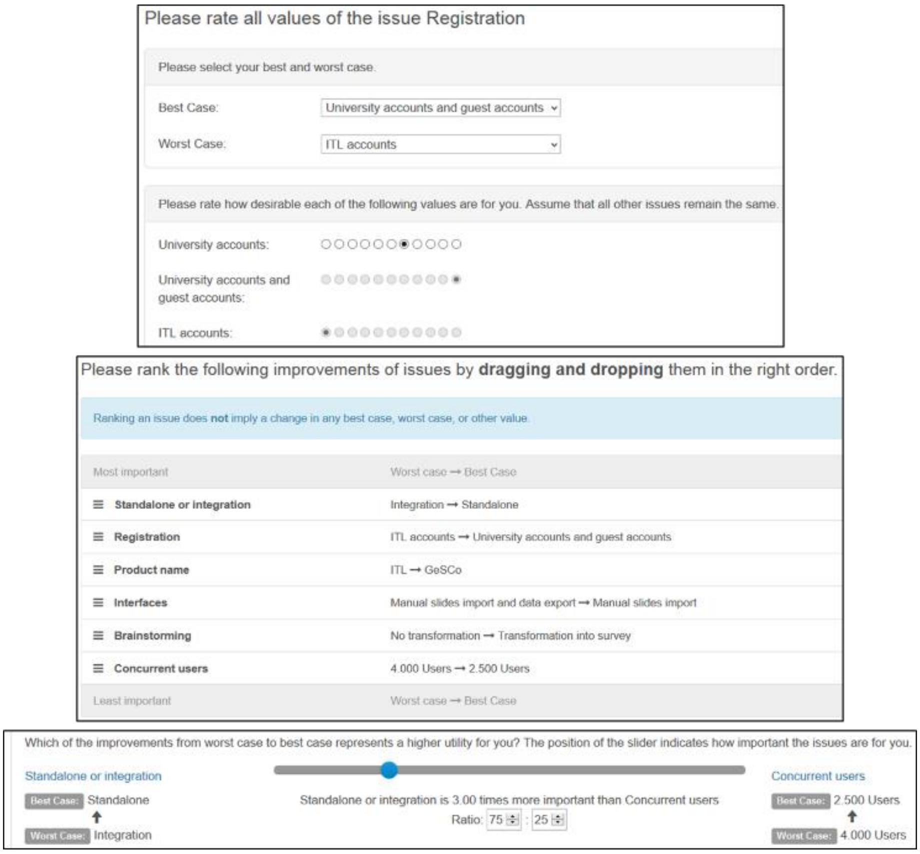


Figure 3. Preference adjustment in Negoisst: Rating of solutions, ranking of requirements, and pairwise comparison of requirements.

5 Scenario-Based Comparison of Dynamic Decision Support for Requirements Negotiation

In the following, we will analyse in a scenario-based [17, 32] comparison as to whether DynaDeS meets its design goals for requirements negotiations. DynaDeS will be compared with decision support based on the WinWin methodology [33] and IntelliReq [31].

5.1 Characteristics of the Decision Support Components to Analyse

The three decision support components all quantify the assessment of requirements and/or solutions. However, they use different assessment bases to provide decision support. In a late incorporation of the WinWin methodology, subsequently called WinWin, TOPSIS is used, which assesses strategic business goals and requirements based on ease of realisation, their business value, and relative penalty [33], while IntelliReq requests the user to rate one of the solutions for a requirement as the preferred solution [31]. Requirements are not assessed. In contrast, DynaDeS uses utility functions to assess requirements and their solutions. It assesses the requirements relatively dependent on all requirements, while the solutions are assessed independently determining one most preferred, one least preferred solution and rating all other solutions directly on a scale from 0 to 1 per requirement.

Regarding the actual support for the negotiators to make their decisions, WinWin focuses on making well-informed decisions based on cost, benefit, and penalty. It additionally considers the requirements' contribution to business goals and performs a relative assessment, so requirements and their solutions can be compared with each other. IntelliReq provides a recommendation for the group of stakeholders, which is based on the preferences of all group members. The stakeholders' assessments do not relate to each other, nor do the assessment of the requirements relate to each other. DynaDeS assesses solutions for requirements under consideration in terms of utility based on the solution preferences and the relative importance of a requirement to facilitate utility analysis [16, 36].

The process for requirements negotiations using the WinWin methodology follows an iterative spiral model. Adjustments can be conducted in each iteration and decision support is correspondingly provided in each iteration when the win conditions are reconciled [6]. IntelliReq is intended to be used in any existing requirements negotiation process model. Since it is not embedded in a process, stakeholders can change their preferences any time as long as the process has not completed resulting in a group recommendation. DynaDeS allows an iterative refinement throughout the negotiation process. Preference adjustments can be conducted at any time to provide updated accurate decision support when asymmetric analysis in the negotiation process is needed.

The concept of requirements and solutions is viewed differently in the three decision support components. The WinWin methodology elicits and negotiates win conditions, issues, and options. To consolidate terminology, we refer to issues as requirements and options as solutions. IntelliReq uses requirements and decision alternatives. Here, we refer to decision alternatives as solutions, too. DynaDeS emerges from general electronic negotiation support. Here, negotiation issues are referred to as requirements and issue values are referred to as solutions.

5.2 Comparison of the Decision Support Components

In the following (see Table 2), we analyse WinWin, IntelliReq, and DynaDeS based on the scenarios from Table 1. Following the WinWin extensions of the spiral model [6], for scenario A, in each iteration a new requirement can be added. The subsequent steps need to be performed, which are to generate solutions to cover the requirement and to jointly agree upon both the new requirement and its solutions. The new requirement has to be assigned to business goals to correctly assess its value. The preference measurement must be repeated. A new solution for a requirement (B) can also be added in each iteration. The solution possibly makes agreements obsolete, so in the same iteration, agreements possibly need to be dissolved and a new agreement must be made involving the new solution. Scenario B does not impact decision support, since solutions are not assessed. The process does not foresee that a stakeholder's preference may change after a requirement is assessed as required in scenario C. A repetition of the assessment and calculation is required. If a stakeholder's preference for a solution (D) has changed, is not applicable in WinWin, since solution preference information is not elicited and included in decision support.

Table 2. Comparison of Decision Support Components for Requirements Negotiations

	Scenario A	*Scenario B*	*Scenario C*	*Scenario D*
WinWin [33]	Repeat preference measurement	No impact on decision support	Repeat assessment for the respective requirement	Not applicable
IntelliReq [31]	Repeat recommendation process	Repeat recommenda-tion process	Not applicable	Repeat recommenda-tion process
DynaDeS	Rate solutions of new requirement; insert requirement into ranking; optionally conduct two requirement comparisons; repeat calculation of utility model	Rate new solution; repeat calculation of utility model	Check requirements ranking; conduct two requirement comparisons; repeat calculation of utility model	Assess solution; repeat calculation of utility model

In IntelliReq, adding a new requirement (A) is not designed. However, adding requirements to the agenda should be possible as long as the superordinate negotiation process is not concluded. For the new requirement solutions have to be entered and all stakeholders must utter their preferences for the best solution. The recommendation process must be repeated from scratch including the new requirement, so decision support can be provided. A new solution (B) is also not foreseen but should be possible to be added at any time. As in scenario A, all stakeholders must consider if the new solution is their most preferred solution and enter it accordingly. The recommendation process must be repeated again. Scenario C is not applicable since requirements are not assessed in IntelliReq. Thus, decision support regarding requirements is not provided. Stakeholders can add or change (D) preferences for solutions at any time. IntelliReq does not use a preference measurement method, but asks for the most preferred solution, while the remaining solutions are not assessed. Thus, changes can be made very easily and quickly. Subsequently the recommendation process must be repeated again.

In DynaDeS, in case of a new requirement (A), all solution preferences can be maintained. Only the solutions of the new requirement have to be assessed. The existing requirements ranking can be reused, too. The ranking is presented to the stakeholder who only must sort in the new requirement. All paired comparisons can be kept. At this point, a utility model can already be calculated, because the relative importance weight of the new requirement can be interpolated based on the requirements ranking. The usual criteria terminate the adjustment process, i.e. two comparisons including the new requirement can be performed if desired. If a new solution is added (B) to an existing requirement, which does not replace the best or worst solution of the respective requirement, it is rated as in scenario A. The partial utility value of the newly rated solution is included in the utility model. In case the preference information for a requirement changes (C), the requirements ranking must be checked respectively corrected and all related paired comparisons must be adjusted. Based thereon, the utility model can be re-calculated. If the preference information of an existing solution has changed (D), the preference adjustment process is the same as in scenario B. The respective requirement and its solution ratings is presented to the stakeholder, who corrects the respective assessment.

6 Discussion and Conclusions

The scenario-based evaluation shows that although both decision support components besides DynaDeS follow dynamic approaches, they do not support preference adjustment reusing the existing assessment. It is required to repeat preference measurement instead of only adjusting changes. Consequentially, the applied preference measurement methods require repeated information input. This is the major advantage of DynaDeS: The analysed decision support component based on the WinWin methodology cannot deal with changes efficiently. Using DynaDeS, the correction of the requirement ranking is already sufficient to measure the relative importance of an issue based on valid relative importance weightings of other requirements. If a more accurate importance weighting is desired, two pairwise comparisons are sufficient.

The main contributions of this paper are the dynamic concept and the proposed decision support process to integrate it. The adjustment of preferences during the requirements negotiation process provides a more accurate basis for decision making that improves the effectiveness and efficiency of the requirements negotiation process. An increase in effectiveness is reached by incorporating new or changing information during the process. Key concepts are both the separation of preferences and utility and relating preferences only to one other object. The general concept can be used and adopted by existing approaches for decision support in requirements negotiations to extend their work. The scenario-based evaluation shows the general applicability of a dynamic perspective as an improvement to decision support. Implications for practice relate to the facilitation of quickly adjusting single preferences, representing the stakeholders' interests, and thus relying on accurate decision support.

Future research could address integrating dependencies between requirements and assumptions of multi-attribute utility theory, since it requires attributes must be independent of each other. Another extension to this work could address an additional level of criteria. Stakeholders may use various criteria to assess requirements or their implementation solutions. For example, customers may evaluate requirements according to their importance, the budget, or the time of delivery, while developers use profitability and availability of resources. Dynamic decision support could enable individual criteria.

Moreover, research proposes that visualisation of information respectively the decision problem should both match the individual's cognitive style and the decision task characteristics [37]. The decision maker's experience could be considered as well [38]. Future studies could investigate, which type of visualisation in requirements negotiation is the most appropriate one. Different types of visualisation and information input could be provided so that an individual decision maker can pick the most appropriate one, e.g. regarding the presentation of the pairwise comparisons in the preference adjustment process. Additionally to different input types, the ratio of the comparison could be provided in numbers, in bars or in a bar chart.

Acknowledgements

The author gratefully acknowledges her PhD scholarship from the State Baden-Württemberg and the support of the research area "Negotiation Research – Transformation, Technology, Media, and Costs", Faculty of Business, Economics, and Social Sciences at the University of Hohenheim.

References

1. Lenz, A., Schoop, M., Herzwurm, G.: Requirements Analysis as a Negotiation Process. In: Kaminski, B., Kersten, G., Szufel, P., Jakubczyk, M., Wachowicz, T. (eds.) The 15th International Conference on Group Decision & Negotiation letters, pp. 303–309. Warsaw School of Economics Press (2015)
2. Grünbacher, P., Seyff, N.: Requirements Negotiation. In: Aurum, A., Wohlin, C. (eds.) Engineering and Managing Software Requirements, pp. 143–162. Springer Berlin Heidelberg (2005)
3. Vetschera, R.: Preference-Based Decision Support in Software Engineering. In: Biffl, S., Aurum, A., Boehm, B., Erdogmus, H., Grünbacher, P. (eds.) Value based software engineering. Springer, Berlin, Heidelberg (2006)
4. Jarke, M., Lyytinen, K.: Editorial: "Complexity of Systems Evolution: Requirements Engineering Perspective". ACM Trans. Manage. Inf. Syst. 5, 1 (2015)
5. Hull, E., Jackson, K., Dick, J.: Requirements engineering. Springer, London (2011)
6. Boehm, B., Kitapci, H.: The WinWin Approach: Using a Requirements Negotiation Tool for Rationale Capture and Use. In: Dutoit, A.H., McCall, R., Mistrík, I., Paech, B. (eds.) Rationale Management in Software Engineering, pp. 173–190. Springer, Berlin, Heidelberg (2006)
7. Ngo-The, A., Ruhe, G.: Decision Support in Requirements Engineering. In: Aurum, A., Wohlin, C. (eds.) Engineering and Managing Software Requirements, pp. 267–286. Springer Berlin Heidelberg (2005)
8. Curhan, J.R., Neale, M.A., Ross, L.: Dynamic valuation. Preference changes in the context of face-to-face negotiation. Journal of Experimental Social Psychology 40, 142–151 (2004)

9. Vetschera, R.: Preference structures and negotiator behavior in electronic negotiations. Decision Support Systems 44, 135–146 (2007)
10. Netzer, O., Toubia, O., Bradlow, E.T., Dahan, E., Evgeniou, T., Feinberg, F.M., Feit, E.M., Hui, S.K., Johnson, J., Liechty, J.C., et al.: Beyond conjoint analysis. Advances in preference measurement. Mark Lett 19, 337–354 (2008)
11. Reiser, A., Krams, B., Schoop, M.: Requirements Negotiation in Consideration of Dynamics and Interactivity. In: Proceedings of the REFSQ 2012 Workshops and the Conference Related Empirical Study, Empirical Fair and Doctoral Symposium, ICB Research Report No. 52, pp. 163–174. Springer, Berlin (2012)
12. Fernandes, M.: Agenda negotiations in electronic negotiation support systems. An information systems perspective. Books on Demand, Norderstedt (2016)
13. Ruhe, G., Eberlein, A., Pfahl, D.: Quantitative WinWin – A New Method for Decision Support in Requirements Negotiation. In: Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering, pp. 159–166 (2002)
14. Reiser, A.: Entscheidungsunterstützung in elektronischen Verhandlungen. Eine Analyse unter besonderer Berücksichtigung von unvollständigen Informationen. Springer Fachmedien Wiesbaden; Imprint: Springer Gabler, Wiesbaden (2013)
15. Lenz, A., Schoop, M., Herzwurm, G.: Electronic Requirements Negotiation - A Literature Survey on the State-of-the-Art. In: UK Academy for Information Systems Conference (ed.) Proceedings of UKAIS 2016, Oxford, paper 23 (2016)
16. Raiffa, H., Richardson, J., Metcalfe, D.: Negotiation analysis. The science and art of collaborative decision making. Belknap Press of Harvard University Press, Cambridge, MA (2002)
17. Pohl, K.: Requirements engineering. Fundamentals, principles, and techniques. Springer, Heidelberg, New York (2010)
18. Carroll, J.M.: Making use. Scenario-based design of human-computer interactions. MIT Press, Cambridge, Mass (2000)
19. Eckert, J., Schaaf, R.: Verfahren zur Präferenzmessung – Eine Übersicht und Beurteilung existierender und möglicher neuer Self-Explicated-Verfahren. Journal für Betriebswirtschaft 59, 31–56 (2009)
20. Netzer, O., Srinivasan, V.: Adaptive Self-Explication of Multiattribute Preferences. Journal of Marketing Research 48, 140–156 (2011)
21. Hwang, C.-L., Lai, Y.-J., Liu, T.-Y.: A new approach for multiple objective decision making. Computers & Operations Research 20, 889–899 (1993)
22. Arnott, D., Pervan, G.: A critical analysis of decision support systems research revisited. The rise of design science. J Inf Technol 29, 269–293 (2014)
23. Keeney, R.L., Raiffa, H.: Decisions with multiple objectives. Preferences and value tradeoffs. Wiley, New York (1976)
24. Kersten, G.E., Noronha, S.J.: WWW-based negotiation support. Design, implementation, and use. Decision Support Systems 25, 135–154 (1999)
25. Thiessen, E.M., Soberg, A.: SmartSettle Described with the Montreal Taxonomy. Group Decis Negot 12, 165–170 (2003)
26. Schoop, M., Jertila, A., List, T.: Negoisst: A Negotiation Support System for Electronic Business-to-Business Negotiations in E-Commerce. Data & Knowledge Engineering 47, 371–401 (2003)
27. Lenz, A., Schoop, M.: Decision Problems in Requirements Negotiations – Identifying the Underlying Structures. In: Schoop, M., Kilgour, D.M. (eds.) Group Decision and Negotiation. A Socio-Technical Perspective: 17th International Conference, GDN 2017,

Stuttgart, Germany, August 14-18, 2017, Proceedings, pp. 120–131. Springer International Publishing, Cham (2017)
28. Herzwurm, G., Schoop, M., Reiser, A., Krams, B.: E-requirements negotiation: Electronic negotiations in the distributed software development. In: Mattfeld, D.C., Robra-Bissantz, S. (eds.) Multikonferenz Wirtschaftsinformatik 2012. Tagungsband der MKWI 2012, pp. 1859–1870. Gito; Univ.-Bibl, Berlin, Braunschweig (2012)
29. Fricker, S., Grünbacher, P.: Negotiation Constellations - Method Selection Framework for Requirements Negotiation. In: Paech, B., Rolland, C. (eds.) Proceedings Requirements Engineering: Foundation for Software Quality. 14th International Working Conference. Springer, Berlin Heidelberg (2008)
30. Schoop, M.: Support of Complex Electronic Negotiations. In: Kilgour, D.M., Eden, C. (eds.) Handbook of Group Decision and Negotiation, 4, pp. 409–423. Springer Netherlands (2010)
31. Felfernig, A., Zehentner, C., Ninaus, G., Grabner, H., Maalej, W., Pagano, D., Weninger, L., Reinfrank, F.: Group Decision Support for Requirements Negotiation. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B. et al. (eds.) Advances in User Modeling, pp. 105–116. Springer, Berlin Heidelberg (2012)
32. Arnott, D., Pervan, G.: Design Science in Decision Support Systems Research. An Assessment using the Hevner, March, Park, and Ram Guidelines. Journal of the Association for Information Systems 13, 923–949 (2012)
33. Kukreja, N., Boehm, B.: Integrating Collaborative Requirements Negotiation and Prioritization Processes: A Match Made in Heaven. In: International Conference on Software and System Process, pp. 141–145 (2013)
34. Grünbacher, P., Köszegi, S., Biffl, S.: Stakeholder Value Proposition Elicitation and Reconsiliation. In: Biffl, S., Aurum, A., Boehm, B., Erdogmus, H., Grünbacher, P. (eds.) Value based software engineering, pp. 133–154. Springer, Berlin Heidelberg (2006)
35. Lenz, A., Schoop, M.: Assessment of Multi-Criteria Preference Measurement Methods for a Dynamic Environment. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 1803-1812 (2019)
36. Bergsmann, J.: Requirements Engineering für die agile Softwareentwicklung. Methoden, Techniken und Strategien. dpunkt, Heidelberg (2014)
37. Engin, A., Vetschera, R.: Information representation in decision making: The impact of cognitive style and depletion effects. Decision Support Systems 103, 94–103 (2017)
38. Gettinger, J., Koeszegi, S.T.: Far from Eye, Far from Heart. Analysis of Graphical Decision Aids in Electronic Negotiation Support. Group Decis Negot 23, 787–817 (2014)

Who is Stressed by Using ICTs? A Qualitative Comparison Analysis with the Big Five Personality Traits to Understand Technostress

Katharina Pflügner, Jens Mattke, and Christian Maier

University of Bamberg, Information Systems and Services, Bamberg, Germany
{katharina.pfluegner,jens.mattke,christian.maier}@uni-bamberg.de

Abstract. The purpose of the current study is to reveal personality profiles that predispose to the experience of techno-stressors within an organizational setting. These insights are useful because techno-stressors lead to considerable costs and adverse health effects. We use the theoretical lens of the transaction-based model of stress to study the effect of the Big Five personality traits on techno-stressors. We distributed a self-rating questionnaire among 221 individuals and analyzed data using fuzzy set Qualitative Comparison Analysis. The results reveal that six different personality profiles lead to the experience of techno-stressors. The study contributes to research by revealing that personality traits need to be investigated in profiles when studying their role in technostress and that different profiles of the Big Five predispose to techno-stressors. The results are useful for practitioners as they allow the prevention of techno-stressors and negative consequences by detecting users who are at risk at an early stage.

Keywords: technostress, personality profiles, Big Five, fuzzy set Qualitative Comparison Analysis (QCA), configurations

1 Introduction

In today's society, people heavily depend on information and communication technology (ICT). Nowadays, living and working without them is hardly conceivable. However, using ICTs can be stressful [1], which is known as technostress [2]. The term was first defined as a disease of adaptation that results from an insufficient dealing with new computer technologies [3]. In current Information Systems (IS) literature, it refers to "*stress that individuals experience due to their use of information systems*" [4, p. 2]. Stress in the workplace leads to substantial costs for organizations and the national economy as well as considerable health impairment for employees. Researchers estimate the annual costs of workplace stress, e.g. by reduced productivity, absenteeism, and compensation, at \$300 billion in the United States, and €20 billion in the EU-15 countries [5]. Technostress depicts an important aspect of stress in the workplace [6]. It is caused by different stimuli, called techno-stressors [2, 7]. In order to diminish financial costs of technostress and adverse consequences like health impairment, it is exceptionally relevant to understand what factors (e.g. personality traits of the user)

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

predispose to the experience of techno-stressors. Such knowledge allows detecting at an early stage, which employees with what personality traits are at risk. Based on this detection, preventive measures and interventions for the reduction of techno-stressors for employees at risk can be implemented even before negative consequences arise.

IS research has mainly focused on evaluating consequences of techno-stressors [2, 7–10] and *environmental or technological characteristics* that precede techno-stressors [1, 10]. The investigation of preceding *characteristics of the user* that increase or decrease techno-stressors has received little attention [4]. This is the case although other research strands beyond technostress inform us that if an individual experiences stressors does not only depend on environmental or technological characteristics, but also characteristics of the individual [11]. Findings suggest that characteristics of the individual include personality traits, which means that personality traits predispose to the level of stressors [12–14]. In line with that, first results from IS research indicate that personality traits are relevant to technostress [8, 9]. The findings reveal that personality traits influence the reaction to techno-stressors and the consequences of techno-stressors [9]. However, the depicted technostress research has two shortcomings: (1) It was not focused on personality traits as predisposing factors to techno-stressors [4] and (2) it has examined personality traits independently, which may only reveal '*half of the picture*' or even lead to false conclusions of the role of personality on techno-stressors [15]. Every user is characterized by various personality traits that coexist, forming a personality profile [16]. Thus, the current study aims to answer the following question: **Which profiles of personality traits predispose to techno-stressors?**

To answer the research question, we base on the transaction-based model of stress [11] and focus on employees working in organizations and experiencing techno-stressors as part of their daily work. Regarding personality traits, we focus on the Big Five personality traits, which reflect higher-level factors [15, 17]. They capture user personality on the most general level and cover most facets of a user's personality. The investigation of Big Five personality traits is analogous to other research in the field of IS [8, 9].

By addressing the research question, we contribute to IS research in the field of technostress and personality. Despite the relevance of technostress and even though we know that personality traits influence the reaction to techno-stressors, personality traits as *predisposing factors* to techno-stressors have not been studied so far in a comprehensive way, referring to premier journals and conferences. Moreover, we highlight the importance of investigating personality profiles within personality research. If a user experiences techno-stressors does not depend on the pure presence of absence of a single personality trait. Instead, it depends on his or her profile of personality traits. We refer to profile of personality traits (equivalent to personality profile) as the specific bundle of the Big Five personality traits. Users have different personality profiles and within each profile, these five personality traits are present to a varying degree. Finally, this research also contributes to practice by revealing personality profiles of how ICTs can evoke techno-stressors and thereby pave the way to reduce the experience of techno-stressors, e.g. through interventions, preventive measures, or sensitization of executives.

2 Theoretical Background

To study how personality profiles predispose to techno-stressors, the transaction-based model of stress [11] guides our research. The theoretical model has been widely used in plenty of different research streams, including technostress literature [2, 4, 18] (for an overview of theories in technostress research see also [19]). Based on that, we introduce the concepts of techno-stressors and personality traits, before depicting the current state of research in these two areas.

2.1 Techno-stressors and Personality in the Transaction-Based Model of Stress

Techno-stressors in the Transaction-Based Model of Stress. The increasing use of ICTs at work as well as daily life expects users to permanently adapt to new functionalities, applications, and work flows. In fact, ICTs act as a new source of stress, which is referred to as technostress [2, 10]. In line with the transaction-based model, *technostress* is a process that results as a combination of a stimulation condition (techno-stressor) and a user's response to the condition (strain). Thus, *technostress* is referred to as the overall process, including techno-stressors, strain, and appraisal as well as coping processes. *Characteristics of the environment* (e.g. organization or technology) and *characteristics of the user* (e.g. personality) precede and influence techno-stressors through appraisal processes [4]. *Techno-stressors* (also called *technostress creators*) are stressful situations caused by ICTs that result in strain [7]. Alleviating factors mitigate the relationship of techno-stressors and strain [4] through coping processes [4, 11, 20]. *Strain* refers to the consequences of techno-stressors. It is the result of being exposed to techno-stressors [10] (see Figure 1).

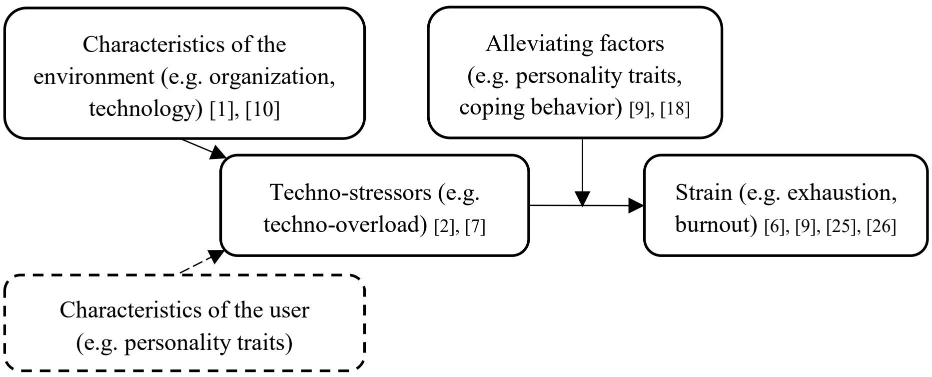


Figure 1. Overview of aspects of technostress and relevant literature. Dotted lines refer to the research gap addressed by the current study.

Personality Traits in the Transaction-Based Model of Stress. Psychological literature informs us that stressors are preceded by characteristics of the individual [11, 20]. Additionally, previous IS research informs us that personality traits as characteristics of the user play an important role in IS contexts as they influence user beliefs and behaviors (for a review see: [21]). The overall tenet thereby posits that personality traits can be captured on different hierarchical levels, whereby these differ regarding their breadth and stability [17]. Broad traits reflect user personality at the highest hierarchical level. At this level, the well-known Big Five personality traits [22] describe elements of an individual's personality. The Big Five personality traits include *neuroticism* (the tendency to experience unpleasant emotions such as anxiety easily), *extraversion* (the tendency to seek the stimulation of others), *openness to experience* (the tendency to prefer new experiences over routines), *conscientiousness* (the tendency to act in a planned and duty-oriented manner), and *agreeableness* (the tendency to cooperate with others) [23]. These five traits are commonly seen as context-free and stable and are useful to understand beliefs and behaviors across different situations [17]. Since the Big Five as broad traits capture user personality on the most general level and cover most facets of a user's personality, we focus on that level.

We have now introduced the concepts of techno-stressors and personality traits and depicted how these are connected. The transaction-based model of stress helps us to understand that personality traits as characteristics of the user precede and influence if an employee experiences techno-stressors. Based on that knowledge, we will now illustrate relevant research findings from the technostress and personality stream. By that, we will reveal research gaps and clarify where the study at hand extends existing research.

2.2 Research Findings Regarding Technostress and Personality

In the following, results from the technostress literature will be explained first, followed by a linkage to personality traits with a focus on IS literature.

Findings Regarding Techno-stressors. Previous research has discussed five techno-stressors. The techno-stressor *techno-overload* describes situations in which users face an increase of work amount and speed due to ICTs. *Techno-invasion* refers to situations where users feel the need to be permanently connected to work and where the line between work and personal life becomes blurred due to ICTs. *Techno-complexity* describes situations where ICT-related complexity leads users to the feeling of inadequate skills and to spend time as well as effort to understand the different aspects of ICTs. *Techno-insecurity* refers to situations where users fear losing their job due to other employees with better ICT skills or due to the replacement by an ICT. Finally, *techno-uncertainty* describes situations where users feel uncertainty because of ongoing changes in ICTs and where they are constantly forced to adapt, learn, and educate themselves about new ICTs [2, 7]. The five mentioned techno-stressors are commonly used in IS research [2, 9, 10].

Findings Regarding Strain. The exposition to techno-stressors leads to increases in strain that are non-beneficial consequences of techno-stressors [10] (for a review see: [4, 24]). In this context, techno-stressors have been linked to adverse job-related consequences like decreases in satisfaction with the ICT, performance, innovation, job satisfaction, and commitment [2, 7, 10]. Additionally, techno-stressors lead to adverse well-being-related consequences like exhaustion, burnout, overall strain, and lower work engagement [1, 6, 8, 9, 18]. These quantitative results are supplemented by qualitative results, suggesting that techno-stressors lead to problems in concentration, sleep, identity, and social relation [25]. It is worth noting that consequences of techno-stressors do not only manifest on a behavioral or emotional level, but also on a biological level [24]. Stress that arises from the use of ICTs even leads to physiological changes. Research in a laboratory experimental setting revealed that the experience of techno-stressors, induced by a system breakdown or high frequency of instant messages, results in endocrinological changes in terms of an incidence in stress hormones [26, 27].

Findings Regarding Preceding and Alleviating Factors. Research has examined technology characteristics that precede techno-stressors, e.g. usefulness, reliability, and presenteeism [1]. Recently, the interest in examining coping strategies that help dealing with techno-stressors has increased. In line with that, aspects of personal control that influence the effect of techno-stressors on strain have been investigated [18].

Findings Regarding User Personality and Technostress. With regards to personality research in IS, research has among others used the Big Five personality traits to show that these influence technological beliefs in terms of perceived usefulness and ease of use [28] as well as user behavior [29]. From the strand of technostress, research informs us about two aspects. First, we see that the Big Five personality traits, such as extraversion and neuroticism, influence how users react to techno-stressors [8]. This means that users high on neuroticism or low on extraversion are more exhausted from using ICTs. Second, we know that the Big Five determine how techno-stressors influence job burnout and job engagement [9]. This means that whether a user is high or low on extraversion determines, among others, if a given level of techno-stressors produces job burnout. In Figure 1, we have summarized the mentioned concepts and have linked them to relevant technostress literature with a focus on studies that have investigated predisposing or alleviating factors to techno-stressors. Moreover, the leverage point of the study at hand is shown dashed.

To sum it up, previous research on technostress emphasizes that there are different techno-stressors that lead to wide-ranging non-beneficial consequences and are preceded by specific technological and organizational characteristics. However, it remains a research gap that no knowledge exists about which personality traits precede whether a user experiences techno-stressors (as shown dashed in Figure 1) (referring to premier journals and conferences). In latest IS literature, there is the claim to investigate how personality traits influence if a user experiences techno-stressors [4] to understand what factors increase the likelihood that ICTs are experienced as stressful. In addition to this research gap and claim, existing studies do not account for the effects of coexisting personality traits, building personality profiles. We know that the Big Five form distinct personality profiles. This suggests that multiple profiles exist, which differ in their personality traits, but all predispose to techno-stressors. Thus, there might be profiles that predispose to techno-stressors, but there might also be profiles of personality traits that do *not* predispose to techno-stressors. Thus, if a user experiences techno-stressors might not depend on the pure presence or absence of a single personality trait. Instead, it might also depend on the presence or absence of other personality traits that coexist and form employees' personality profile. Yet, so far, research has not considered which different personality profiles predispose to techno-stressors. This knowledge is essentially important to prevent, intervene, and foster the handling of techno-stressors.

To address the mentioned research gaps, we apply a configurational approach using fuzzy set Qualitative Comparison Analysis (fsQCA) [30]. Previous research on user personality emphasizes that there are three different hierarchical levels. Thereby, it is recommended to initially focus on broad traits, particularly if no previous research exists that identifies other, more specific, narrower traits [31]. Therefore, we focus on the Big Five personality traits [16] to explain how personality profiles predispose to experiences of techno-stressors. Having in mind that techno-stressors result in adverse consequences like decreased performance and IT use, the results allow the drawing of propositions to positively react to techno-stressors and related adverse effects.

3 Method

We next describe the data collection process as well as the used measures, and outline the validity and reliability of the measurement model. We then explain our data analysis using fuzzy set Qualitative Comparison Analysis (fsQCA) [30], which has been successfully used in IS research to examine different personalities using individual level data [32, 33].We use this approach to reveal which configurations of different personality traits lead to the same outcome (techno-stressors). Here, configurations are equivalent to profiles of the Big Five personality traits. Thus, each configuration encompasses the five personality traits that are pronounced with varying degree.

Data Collection. The sampling strategy is to get a broad spectrum of participants who are familiar to work with IS on a regular basis as their profession. Therefore, we chose to invite cloud workers using Amazon Mechanical Turk (mTurk). Those participants professionally work with IS and mostly have additional occupations in organizations [34]. We prepared an online survey and used mTurk, which has become an established approach in IS research and is equal to traditional data collection approaches [35]. To ensure a high quality of our data, we embedded two attention tests in our survey. Overall, 239 individuals participated, however, we removed 18 participants as they failed the attention tests. The final sample consists of 221 participants. Thus, the sample size is large enough, as the ratio of conditions (here: the Big Five) to number of participants should be smaller than 0.2 and in this study we have a ratio of 0.022. This means fsQCA only requires a sample of 25 observations, yet our sample size is eight times as large [36]. Individuals who successfully participated in the survey received $ 0.20. The characteristics of the final sample are shown in Table 1.

Table 1. Sample characteristics (in per cent) of 221 participants

Age (in years)	< 20	1.0	ICT use (hours per week)	< 10	10.9
	20-29	40.1		10-19	12.5
	30-39	40.7		20-29	16.2
	40-49	12.5		30 – 39	16.1
	> 49	5.7		> 39	44.3
Sex	Female	39.1	IT professional	No	64.6
	Male	60.9		Yes	35.4

We followed recommendations for self-reported data [37] and tested for common method bias. For this, we applied Harman's single factor test, which reveals that one factor only explains 37 percent of the variance, which is below the recommended threshold of 50 percent [38]. Additionally, we examined the correlation matrix (see Table 4), which does not reveal any high correlations [39]. In summary, we can state that common method bias is not an issue in this study.

Measures. To measure the Big Five personality traits, we base on existing items [9]. For openness to experience we used three items, yet one item was removed because the loading was below the recommended threshold [40]. For neuroticism, we used three items and for agreeableness three items, but one item was removed due to a low loading. To measure conscientiousness and extraversion, we used three items each [9]. To measure techno-stressors, we used the five techno-stressors and measured overall techno-stressors as a second-order construct resulting of the five stressors [2]. Therefore, we base all techno-stressor items on Ragu-Nathan [2]. We measured techno-overload by five items, but we removed one item because of a low loading. We used four items to measure techno-complexity and again removed one item due to a low loading. We used five items for techno-complexity, four for techno insecurity, and four for techno-uncertainty. All items were measured on a 7-point Likert-type agreement scale, ranging from 1 (completely disagree) to 7 (completely agree) (see Table 2).

Table 2. Measures

Techno-overload [2], $\alpha = 0.86$	I am forced by ICTs to work much faster. [0.795] I am forced by ICTs to do more work than I can handle. [0.882] I am forced by ICTs to work with very tight time schedules. [0.881] I have a higher workload because of increased ICT complexity. [0.743]
Techno-invasion [2], $\alpha = 0.79$	I have to be in touch with my work even during my vacation due to ICTs. [0.896] I have to sacrifice my vacation and weekend time to keep current on new ICTs. [0.900] I feel my personal life is being invaded by ICTs. [0.720]
Techno-complexity [2], $\alpha = 0.86$	I do not know enough about ICTs to handle my job satisfactorily. [0.721] I need a long time to understand and use new ICTs. [0.962] I do not find enough time to study and upgrade my ICT skills. [0.741] I find new recruits to this organization know more about ICTs than I do. [0.719] I often find it too complex for me to understand and use new ICTs. [0.787]
Techno-insecurity [2], $\alpha = 0.86$	I feel constant threat to my job security due to new ICTs. [0.853] I am threatened by coworkers with newer ICT skills. [0.891] I do not share my knowledge with my coworkers for fear of being replaced. [0.800] I feel there is less sharing of knowledge among coworkers for fear of being replaced. [0.725]
Techno-uncertainty [2], $\alpha = 0.83$	There are always new developments in the technologies we use in our organization. [0.745] There are constant changes in computer software in our organization. [0.846] There are constant changes in computer hardware in our organization. [0.803] There are frequent upgrades in computer networks in our organization. [0.852]

Openness [9], α = 0.87	I see myself as creative. [0.961] I see myself as imaginative. [0.921]
Neuro-ticism [9], α = 0.75	I see myself as moody. [0.799] I see myself as easily upset. [0.846] I see myself as anxious. [0.817]
Agreeable-ness [9], α = 0.75	I see myself as sympathetic. [0.728] I see myself as kind. [0.991]
Conscien-tiousness [9], α = 0.74	I see myself as dependable. [0.761] I see myself as self-disciplined. [0.714] I see myself as organized. [0.909]
Extraver-sion [9], α = 0.85	I see myself as extraverted. [0.894] I see myself as enthusiastic. [0.851] I see myself as talkative. [0.880]

Measurement Model. To ensure content validity we only used items that have been used and validated in previous research. Each item used in this study has a loading above 0.707, which attests indicator reliability [2]. We can attest construct reliability because the average variance extracted (AVE) of each construct is higher than 0.50 and the composite reliability (CR) is higher than 0.70. Furthermore, we can attest discriminant validity, as we conducted the heterotrait-monotrait (HTMT) ratio, which is 0.68 and consequently below the threshold of 0.85 [41]. The square root of the AVE is higher than the corresponding correlations of the constructs [42, 43] (see Table 4).

Data Analysis Using fsQCA. To analyze which configurations of the five personality traits predispose to techno-stressors, we take a configurational approach [44]. More precisely, we use fuzzy set Qualitative Comparison Analysis (fsQCA) [30], which enables us to study techno-stressors as the result of a configuration of personality traits. In this study, a configuration refers to a specific bundle of the five personality traits which can predispose to techno-stressors and where each personality trait is expressed through a fuzzy set. Using fuzzy sets allows us to express the degree to which a measure belongs to a personality trait.

The *data analysis of configurations sufficient for techno-stressors* consists of four subsequent steps [30, 45]. First, for the *calibration* of the survey data into fuzzy sets, we applied the direct calibration, as recommended in QCA literature [44, 45]. For this, we used three recommended qualitative anchors (value 1 for full non-membership, value 4 for the crossover point, and value 7 for full membership) and calibrated the survey data [44, 45]. The resulting fuzzy sets are used in the subsequent steps of the analysis. Second, based on the fuzzy sets, we perform the *construction of the truth table*, which lists all possible configurations of the five personality traits. Third, after constructing the truth table, we applied the recommended thresholds to reduce the truth table to *meaningful configurations* [46]. In line with previous research [45, 47], we applied a frequency threshold of three, meaning that all configurations with less than three observations are dropped from further analysis. In conformity with QCA literature [46], we applied a *consistency threshold* of 0.90, meaning that only configurations with a consistency of at least 0.90 are considered in the analysis. Consistency is a measure that captures the extent to which a configuration leads to the studied outcome [30]. This means that the extent to which a configuration of personality traits predisposes to the sum of techno-stressors is measured. By applying those two thresholds, we reveal configurations that are sufficient in bringing about techno-stressors [45]. Sufficient means that every time the configuration of these personality traits is present, the experience of techno-stressors is present as well.

4 Results

In this section, we will outline the findings of the analysis of configurations sufficient for techno-stressors.

Six Different Configurations of Personality Traits Predispose to the Experience of Techno-stressors. Results reveal six alternative configurations of personality traits sufficient for predisposing to techno-stressors. We draw on graphical illustration of the configurations for readability reasons (see Table 3).

Table 3. Configurations of personality traits predisposing to the experience of techno-stressors

Configuration / Conditions	*The neurotic-agreeable personality profile* *C1*	*The neurotic-conscientious personality profile* *C2*	*The agreeable-conscientious personality profile* *C3*	*The neurotic-agreeable and conscientious personality profile* *C4*	*The open-extraverted and agreeable personality profile* *C5*	*The open-neurotic personality profile* *C6*
Openness to experience			⊗	⊗	●	●
Neuroticim	●	●		●	⊗	●
Agreeableness	●	⊗	●	●	●	⊗
Conscientiousness	⊗	●	●	●	⊗	
Extraversion	⊗	⊗	⊗		●	⊗
Raw coverage	0.38	0.36	0.36	0.37	0.37	0.36
Unique coverage	0.03	0.01	0.02	0.02	0.03	0.01
Consistency	0.90	0.90	0.90	0.90	0.91	0.91
Solution coverage	0. 60					
Solution consistency	0. 81					

Note: Black circles (●) indicate the presence of a personality trait, white crossed-out circles (⊗) the absence of a personality trait, and blank spaces indicate a *don't care* situation. In this case, the trait plays a subordinate role and may be either present or absent. C means configuration.

The configurations are translated in the following way: Black circles indicate the presence of a personality trait, white crossed-out circles the absence of a personality trait, and blank spaces indicate a '*don't care*' situation. In this case, the trait plays a subordinate role and may be either present or absent. The configurations (C1 – C6) are named based on their main characteristics, e.g. neurotic-agreeable personality profile (C1), as users with this configuration show high levels of neuroticism and agreeableness. In summary, we identified the following configurations (personality profiles): Neurotic-agreeable (C1), neurotic-conscientious (C2), agreeable-conscientious (C3), neurotic-agreeable and conscientious (C4), open-extraverted and agreeable (C5), and open-neurotic (C6).

The solution coverage of 0.60 indicates the degree of how much of the outcome (here: techno-stressors) is covered by the six configurations. Thus, the six configurations account for 60 percent of the membership in the outcome, which illustrates a high explanatory power of the configurations [48]. In line with recommendations in QCA literature [46], the solution consistency of 0.81 as well as the consistency of each configuration exceed the minimum value of 0.80. The raw coverage of the six configurations expresses the "*proportion of membership in the outcome explained by each term of the solution*" [49, p. 86], which means the extent to which the configuration covers the cases of the outcome, and ranges from 0.36 to 0.38. The six unique coverage values range from 0.01 to 0.03, expressing the unique contribution of each configuration under exclusion of the contribution of other configurations [30].
Each of the Six Configurations that were Identified Substantially and Equally Contribute to Experiencing Techno-stressors. Comparing the six configurations, the raw coverage scores and unique scores are comparable, meaning that the configurations explain techno-stressors to an equivalent extent. Furthermore, the configurations have comparable relative importance for explaining techno-stressors. To further assess the mentioned results, we discuss them in the following section, providing theoretical and practical relevance as well as future research directions.

Table 4. Descriptive statistics

Constructs	M	SD	CR	AVE	1	2	3	4	5	6	7	8	9	10
1 Open	5.39	1.34	0.94	0.89	0.94									
2 Neuro	3.95	1.32	0.86	0.67	-0.06	0.82								
3 Agree	5.43	1.11	0.85	0.74	0.19	-0.17	0.85							
4 Cons	5.29	1.21	0.83	0.63	0.16	-0.14	0.09	0.79						
5 Extra	4.28	1.47	0.91	0.76	0.27	-0.16	0.32	0.16	0.87					
6 TO	3.88	3.88	0.90	0.70	0.06	0.20	0.06	-0.04	0.10	0.84				
7 TInv	3.97	3.96	0.87	0.69	0.17	0.10	0.05	0.08	0.11	0.60	0.83			
8 TC	3.12	3.12	0.89	0.63	-0.10	0.09	-0.02	-0.02	0.15	0.35	0.28	0.79		
9 TU	3.08	3.10	0.89	0.66	0.22	0.13	-0.08	0.10	0.13	0.29	0.37	0.23	0.81	
10 TIns	4.73	4.73	0.89	0.61	-0.09	0.19	-0.04	-0.04	0.04	0.51	0.41	0.65	0.28	0.78

Note: square root of AVE is listed on the diagonal of bivariate correlations; techno-stressors are calculated as the sum of constructs 6 to 10, Open = openness, Neuro = neuroticism, Agree = agreeableness, Cons = conscientiousness, Extra = extraversion, TO = techno-overload, TInv = techno-invasion, TC = techno-complexity, TU = techno-uncertainty, TIns = techno-insecurity, M = mean, SD = standard deviation; CR = composite reliability; AVE = average variance extracted

5 Discussion

Using information and communication technologies (ICTs) can lead to the experience of techno-stressors, which in turn results in negative consequences for organizations, e.g. reduced productivity, and for users, e.g. emotional exhaustion [7, 9]. For both sides, knowledge about what factors predispose to this type of stress is relevant to address and deal with it. The aim of the paper at hand is to provide a better understanding of the interplay of personality traits influencing users' experience of techno-stressors. We base on the transaction-based model of stress [11, 20] to explain that personality profiles influence the level of techno-stressors that employees experience, which in turn influences health and organizational consequences [9, 10, 26]. To answer the research question, we test for different personality profiles leading to techno-stressors, and discovered six profiles of the Big Five personality traits by the application of fsQCA. In the following, we outline the theoretical and practical implications followed by limitations and possible areas of future research.

Theoretical Contributions. The current research contributes to the two research strands of technostress and personality as illustrated in the following.

Up to now, personality traits as predispositions to techno-stressors have gained little attention (see Figure 1). Personality traits are characteristics of the user. Existing literature revealed personality traits to directly influence the consequences of techno-stressors, e.g. exhaustion [8], and to moderate the relationship between techno-stressors and consequences like burnout and job engagement [9].

Thus, the study at hand adds the importance of personality traits, especially profiles of personality traits, as predispositions to techno-stressors to existing literature (as shown dashed in Figure 1). The findings illustrate that there is not one single personality profile, which leads to techno-stressors. Instead, there are multiple profiles leading to the experience of techno-stressors. We see that neuroticism plays a major role in four personality profiles (C1, C2, C4, C6). However, in two of the profiles (C3, C5) neuroticism plays a subordinate role or must even be absent. We see that being neurotic is not the only way towards experiencing techno-stressors. In the open-extraverted and agreeable personality profile (C5), neuroticism even has a negative influence on techno-stressors. Similarly, four of the personality profiles encompass the absence of extraversion (C1, C2, C3, C6). However, in two of the profiles (C4, C5) extraversion does not have any influence at all or can even have a positive influence on techno-stressors. Based on that, we see that there are different profiles of personality traits that lead to the experience of techno-stressors and that they explain techno-stressors to an equivalent extent. It is highlighted that certain personality traits do not necessarily have a positive or negative influence and that it rather depends on the combination of traits if they predispose to techno-stressors.

Practical Contributions. This research contributes to practice as the findings enable the early detection of users who are at risk, even before negative consequences arise. The findings can be used for preventing that employees are stressed by using ICTs. This is relevant based on the fact that techno-stressors negatively affect employees' health and organizational factors [7, 9, 26, 27].

Who might incorporate and contribute from the research findings? First, employees themselves may use the findings for self-insight. They may evaluate themselves based on the depicted profiles of personality traits, which allows them to identify if they are at risk. This self-insight is the first, but very important step to take action for mitigating and preventing techno-stressors. Based on that self-insight, they may e.g. change their ICT work habits, seek support, or engage in stress management practices.

Second, the depicted research findings may sensitize leaders for the risk of techno-stressors. More concretely, the findings highlight that employees differ in their risk of experiencing techno-stressors. There are employees that are especially prone to experience techno-stressors, whereas other employees are not. Thus, leaders should be aware of these differences, which should also be reflected in their leadership behavior. At this point, a scattergun approach might not be helpful. Instead, the findings allow leaders to specifically focus on employees at risk, which might include individual support provision or the provision of stress management interventions for the risk group.

Third, the findings can be used by software designers for the installation of adaptive enterprise systems. In combination with information from a voluntary personality screening, techno-stressors may be prevented specifically for the employees who are at risk by automatic technological adaptations, e.g. temporary interruption of mail servers for a reduction of incoming mails or the application of wizards [50].

Limitations and Further Research. There are some limitations to the results provided in this study, as illustrated in the following. To begin with, we focused on investigating user's personality traits on the most general level in a way that covers most facets of personality (Big Five) to understand which profiles of personality traits predispose to techno-stressors. However, investigating personality profiles of narrower traits, e.g. IT mindfulness [17], also seems reasonable and could be focus of future research as users can change these personality traits to some degree [17]. Moreover, we focused on techno-stressors as an aggregated construct and did not investigate the predisposition to personality profiles for each techno-stressor separately. Here, future research might study whether each techno-stressor is influenced by different profiles. Finally, it was not specifically investigated how demographic or organizational factors like type of organization or participant's role within the organization influence how personality traits predispose to techno-stressors as this was the first approach to gain an understanding of the depicted connection. The Big Five personality traits are stable across situations and time [51, 52]. Thus, they are rather independent from the influence of organizational factors. However, future research could take a look deeper into each of the six personality profiles and their predisposition to techno-stressors to reveal possible differences based on demographic or organizational factors.

6 Conclusion

The invasion of ICTs at work and daily life is constantly growing, and with it techno-stressors and related financial costs as well as negative health consequences. Therefore, knowledge is needed to detect users who are at risk at an early stage. Using a QCA approach, the study identifies six different profiles of personality traits to predispose to techno-stressors. These findings contribute to existing research by highlighting the importance of personality profiles as a predisposition to techno-stressors and the need to investigate personality traits in profiles. Practitioners may use the research findings to intervene and prevent negative consequences arising from techno-stressors.

References

1. Ayyagari, R., Grover, V., Purvis, R.: Technostress: Technological antecedents and implications. MIS Quarterly 35, 831-858 (2011)
2. Ragu-Nathan, T.S., Tarafdar, M., Ragu-Nathan, B.S., Tu, Q.: The consequences of technostress for end users in organizations: Conceptual development and empirical validation. Information Systems Research 19, 417–433 (2008)
3. Brod, C.: Technostress. The human cost of the computer revolution. Addison Wesley Publishing Company (1984)
4. Tarafdar, M., Cooper, C.L., Stich, J.-F.: The technostress trifecta - techno eustress, techno distress and design: Theoretical directions and an agenda for research. Information Systems Journal 9, 1–37 (2017)
5. European Agency for Safety and Health at Work (EU-OSHA): Calculating the cost of work-related stress and psychosocial risks: European risk observatory literature review. Publications Office of the European Union, Luxembourg (2014)
6. Maier, C., Laumer, S., Eckhardt, A.: Information technology as daily stressor: Pinning down the causes of burnout. Journal of Business Economics 85, 349–387 (2015)
7. Tarafdar, M., Tu, Q., Ragu-Nathan, B.S., Ragu-Nathan, T.S.: The impact of technostress on role stress and productivity. Journal of Management Information Systems 24, 301–328 (2007)
8. Maier, C., Laumer, S., Weinert, C., Weitzel, T.: The effects of technostress and switching stress on discontinued use of social networking services: A study of Facebook use. Information Systems Journal 25, 275–308 (2015)
9. Srivastava, S.C., Chandra, S., Shirish, A.: Technostress creators and job outcomes: Theorising the moderating influence of personality traits. Information Systems Journal 25, 355–401 (2015)
10. Tarafdar, M., Tu, Q., Ragu-Nathan, T.S.: Impact of technostress on end-user satisfaction and performance. Journal of Management Information Systems 27, 303–334 (2011)
11. Lazarus, R.S., Folkman, S.: Stress, appraisal, and coping. Springer, New York (1984)
12. Gunthert, K.C., Cohen, L.H., Armeli, S.: The role of neuroticism in daily stress and coping. Journal of Personality and Social Psychology 77, 1087–1100 (1999)
13. Schneider, T.R., Rench, T.A., Lyons, J.B., Riffle, R.R.: The influence of neuroticism, extraversion and openness on stress responses. Stress & Health 28, 102–110 (2012)
14. Spector, P.E., O'Connell, B.J.: The contribution of personality traits, negative affectivity, locus of control and Type A to the subsequent reports of job stressors and job strains. Journal of Occupational and Organizational Psychology 67, 1–12 (1994)
15. Grant, S., Langan-Fox, J.: Occupational stress, coping and strain: The combined/interactive effect of the Big Five traits. Personality and Individual Differences 41, 719–732 (2006)
16. McCrae, R.R., Costa, P.T.: Personality in adulthood: A five-factor theory perspective. Guilford Press, New York (2006)
17. Thatcher, J.B., Wright, R.T., Sun, H., Zagenczyk, T.J., Klein, R.: Mindfulness in information technology use: Definitions, distinctions, and a new measure. MIS Quarterly 42, 831-847 (2018)
18. Galluch, P.S., Grover, V., Thatcher, J.: Interrupting the workplace: Examining stressors in an information technology context. Journal of the Association for Information Systems 16, 1–47 (2015)
19. Fischer, T., Riedl, R.: Theorizing technostress in organizations. A cybernetic approach. Proceedings of the 12th International Conference on Wirtschaftsinformatik (2015)

20. Cooper, C.L., Dewe, P.J., O'Driscoll, M.P.: Organizational stress: A review and critique of theory, research, and applications. Sage Publications (2001)
21. Maier, C.: Personality within information systems research: A literature analysis. Proceedings of the 20th European Conference on Information Systems (ECIS), Barcelona (2012)
22. Costa, P.T., McCrae, R.R.: Set like plaster? Evidence for the stability of adult personality. In: Heatherton, T.F., Weinberger, J.L. (eds.) Can personality change?, pp. 21–40. American Psychological Association, Washington DC (1997)
23. Goldberg, L.R.: Language and individual differences: The search for universals in personality lexicons. Review of Personality & Social Psychology 2, 141–165 (1981)
24. Riedl, R.: On the biology of technostress. Literature review and research agenda. DATA BASE for Advances in Information Systems 44, 18–55 (2013)
25. Salo, M., Pirkkalainen, H., Koskelainen, T.: Technostress and social networking services: Explaining users' concentration, sleep, identity, and social relation problems. Information Systems Journal 40, 182 (2018)
26. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress from a neurobiological perspective. System breakdown increases the stress hormone cortisol in computer users. Business & Information Systems Engineering 4, 61–69 (2012)
27. Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., Grover, V.: NeuroIS - alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. Journal of the Association for Information Systems 15, 723–753 (2014)
28. Devaraj, S., Easley, R.F., Crant, J.M.: How does personality matter? Relating the five-factor model to technology acceptance and use. Information Systems Research 19, 93–105 (2008)
29. McElroy, J.C., Hendrickson, A.R., Townsend, A.M., DeMarie, S.M.: Dispositional factors in internet use: Personality versus cognitive style. MIS Quarterly 31, 809–820 (2007)
30. Ragin, C.C.: The comparative method. Moving beyond qualitative and quantitative strategies. University of California Press (2014)
31. Eckhardt, A., Laumer, S., Maier, C., Weitzel, T.: The effect of personality on IT personnel's job-related attitudes: establishing a dispositional model of turnover intention across IT job types. Journal of Information Technology 31, 48–66 (2016)
32. Mattke, J., Müller, L., Maier, C.: Why do individuals avoid social media advertising. A qualitative comparison analysis study. Proceedings of the 26th European Conference on Information Systems (ECIS) (2018)
33. Mattke, J., Müller, L., Maier, C., Weitzel, T.: Bitcoin resistance behavior. A QCA study explaining why individuals rests bitcoin as a means of payment. Proceedings of the International Conference on Information Systems (ICIS) (2018)
34. Ross, J., Zaldivar, A., Irani, L., Tomlinson, B.: Who are the crowdworkers? Shifting demographics in amazon mechanical turk. CHI EA, 2863–2872 (2010)
35. Steelman, Z.R., Hammer, B.I., Limayem, M.: Data collection in the digital age. Innovative alternatives to student samples. MIS Quarterly 38, A1-A20 (2014)
36. Marx, A.: Towards more robust model specification in QCA results from a methodological experiment. American Sociological Association, Philadelphia, PA (2006)
37. Chin, W.W., Thatcher, J.B., Wright, R.T.: Assessing common method bias. Problems with the ULMC technique. MIS Quarterly 36, 1003–1019 (2012)
38. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research. A critical review of the literature and recommended remedies. Journal of Applied Psychology 88, 879–903 (2003)

39. Pavlou, P., Liang, H., Xue, Y.: Understanding and mitigating uncertainty in online exchange relationships. A principle-agent perspective. MIS Quarterly 31, 105–136 (2007)
40. Carmines, E.G., Zeller, R.A.: Reliability and validity assessment. Sage Publ., Newbury Park, Calif. (2008)
41. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science 43 (2014)
42. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18, 39–50 (1981)
43. Hulland, J.S.: Use of partial least squares (PLS) in strategic management research: A review of four recent studies. Strategic Management Journal 20, 195–204 (1999)
44. Misangyi, V.F., Greckhamer, T., Furnari, S., Fiss, P.C., Crilly, D., Aguilera, R.: Embracing causal complexity. The emergence of a neo-configurational perspective. Journal of Management 43, 255–282 (2017)
45. Fiss, P.C.: Building better causal theories. A fuzzy set approach to typologies in organization research. Academy of Management Journal 54, 393–420 (2011)
46. Schneider, C.Q., Wagemann, C.: Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. Comparative Sociology 9, 397–418 (2010)
47. Müller, L., Mattke, J., Maier, C., Weitzel, T.: The curse of mobile marketing: A mixed methods study on individuals' switch to mobile ad blockers. Proceedings of the International Conference on Information Systems (ICIS) (2017)
48. Liu, Y., Mezei, J., Kostakos, V., Li, H.: Applying configurational analysis to IS behavioural research: A methodological alternative for modelling combinatorial complexities. Information Systems Journal 27, 59–89 (2017)
49. Ragin, C.C.: User's Guide to Fuzzy-Set/Qualitative Comparative Analysis 2.0. Department of Sociology, University of Arizona, Tucson, Arizona (2006)
50. Adam, M.T.P., Gimpel, H., Maedche, A., Riedl, R.: Design blueprint for stress-sensitive adaptive enterprise systems. Business & Information Systems Engineering 59, 277–291 (2017)
51. Cobb-Clark, D., Schurer, S.: The stability of big-five personality traits. IZA Discussion Papers 5943 (2011)
52. Soldz, S., Vaillant, G.E.: The Big Five personality traits and the life course: A 45-year longitudinal study. Journal of Research in Personality 33, 208–232 (1999)

Walking the Middle Path: How Medium Trade-Off Exposure Leads to Higher Consumer Satisfaction in Recommender Agents

Veronika Schuhbeck[1], Nils Siegfried[2], Verena Dorner[3], Alexander Benlian[2], Michael Scholz[1], Guido Schryen[4]

[1]University of Passau, Institute of Information Systems,
Innstraße 41, 94032 Passau, Germany
veronikaschuhbeck@yahoo.de, michael.scholz@uni-passau.de

[2]Darmstadt University of Technology, Institute of Systems & E-Services
Hochschulstraße 1, 64289 Darmstadt, Germany
siegfried@ise.tu-darmstadt.de, benlian@ise.tu-darmstadt.de

[3]Karlsruher Institute of Technology,
Institute of Information Systems and Marketing,
Fritz-Erler-Straße 23, 76133 Karlsruhe, Germany
verena.dorner@kit.edu

Institute of Management Informtation Systems and Operations Research,
Paderborn University, Germany
Warburger Straße 100, 33098 Paderborn
guido.schryen@uni-paderborn.de

Abstract. Recommender Agents (RAs) facilitate consumers' online purchase decisions for complex, multi-attribute products. As not all combinations of attribute levels can be obtained, users are forced into trade-offs. The exposure of trade-offs in a RA has been found to affect consumers' perceptions. However, little is known about how different preference elicitation methods in RAs affect consumers by varying degrees of trade-off exposure. We propose a research model that investigates how different levels of trade-off exposure cognitively and affectively influence consumers' satisfaction with RAs. We operationalize these levels in three different RA types and test our hypotheses in a laboratory experiment with 116 participants. Our results indicate that with increasing trade-off exposure, perceived enjoyment and perceived control follow an inverted U-shaped relationship. Hence, RAs using preference elicitation methods with medium trade-off exposure yield highest consumer satisfaction. This contributes to the understanding of trade-offs in RAs and provides valuable implications to e-commerce practitioners.

Keywords: Recommender Agents, Preference Elicitation Method, Trade-off Exposure, Customer Satisfaction

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

1 Introduction

Demand for products customized to individual consumers' preferences (e.g., mass customized or custom-made) has grown hugely in recent years. Concurrently, the range of product variations available on the market has expanded [1]. As a result, consumers are increasingly in need of decision support systems that help them sift through the many variations by providing information transparency on product attributes (e.g., price, quality, performance) and their interrelationships (e.g., price-quality trade-offs). Prior research indicates that navigation aids and decision aids such as recommendation agents (RAs) help consumers understand their preferences and find suitable products [2]. A central aspect within RAs, which is still subject to discussion, are trade-offs.

Trade-offs are mutual dependencies between two attributes, where certain attribute levels become unavailable (e.g. low price) as a result of having selected a specific level of another attribute (e.g. high quality) [3; 4]. How trade-offs are integrated and presented in RAs is an important issue since trade-offs can be difficult to process and imply losses in certain attributes, which tends to make consumers unhappy and frustrated – and in turn reduces their willingness to use the RA [5].

Recommender agents use a number of different strategies to elicit consumer preferences and, hence, present trade-offs differently [6]. Trade-off exposure reflects the degree to which the consumer is forced to recognize and deal with trade-offs. In other words, trade-off exposure refers to how much the user is forced to think about availability and unavailability of product attributes dependent on prior choices in other attributes. For example, when choosing a specific level of quality automatically rules out a variety of choice options (e.g., cheap alternatives), the user is confronted with high trade-off exposure, while an RA with low trade-off exposure hides such consequences. Lower levels of exposure reduce decision difficulty but have the drawback of depriving consumers of important information [7]. In contrast, higher levels of exposure increase recommendation accuracy but tend to be cognitively more challenging for consumers [8].

While traditional RA designs have either hidden trade-offs from users or forced them to explicitly compare attribute levels, recent designs exhibit a medium level of trade-off exposure [5; 9]. While the level of trade-off exposure in an RA may significantly affect users' perceptions (e.g. effort of using or satisfaction with the result), its impact is not yet fully understood. Xu et al. [5] have found that a medium level of trade-off exposure leads to highest usage intentions, however, they have only varied the number of trade-offs while using the same RA in all conditions. Hence, we are investigating the follow research question:

How do preference elicitation methods with different levels of trade-off exposure affect consumers' satisfaction with an RA?

Our study contributes to understanding the role of trade-off exposure by comparing RAs of low, medium and high trade-off exposure in a laboratory experiment with 116 participants. To investigate the influence of trade-off exposure on satisfaction with an RA, we examine cognitive and affective perceptions [10], and how they in turn affect decision quality and decision effort, preceding consumers' satisfaction with the RA.

The results of our study provide valuable contributions to theory and practice. From a theory perspective, we contribute to the understanding of how medium trade-off exposure affects consumer perceptions and, ultimately, how it affects consumers' satisfaction with RAs. From a practitioner's perspective, this study helps RA designers to facilitate consumers' purchase processes. It advances their understanding of how the level of trade-off exposure affects consumers and which level to choose when designing an RA. This in turn can lead to increased sales as consumers find the right products more easily among an extensive range of product variations.

2 Theoretical Background

Recommender agents are virtual assistants which aim to help consumers find, with as little effort as possible, the product that fits their requirements best [11]. The two most common approaches for RAs, collaborative filtering and content-based recommender systems, use information about consumers' past purchases to predict future purchase decisions [12]. However, due to lack of data for new consumers and new products (the so-called cold-start problem), recommendation quality may be quite low [9]. Systems that do not suffer from these problems are based on multi-attribute value theory (MAVT). MAVT-based RAs use consumer-specific value functions and attribute importance weights, estimated at the time of purchase, to provide recommendations [13].

How well an RA is able to support consumer decisions depends, among other factors, on how it deals with trade-offs. Trade-offs are mutual dependencies between two attributes, where certain attribute levels become unavailable as a result of having selected a specific level of another attribute [3; 4]. The trade-offs can be shown explicitly to the consumer or implicitly which is referred to as trade-off exposure. RAs with high trade-off exposure explicitly inform consumers when certain attribute levels become unavailable as a result of having selected a specific level of another attribute. Unavailable attribute combinations are presented to consumers and require consumers to take corrective action. RAs with low trade-off exposure hide consumers from this information, which means unavailable attributes are not shown or attribute selection is decoupled from product presentation. Hence, RAs with high trade-off exposure generally provide users with a more transparent (i.e., users can inspect choice consequences) but also more challenging decision process.

The level of trade-off exposure depends largely on the preference elicitation method implemented in a MAVT-based RA. Common preference elicitation methods are absolute measurement, pairwise comparisons, ordinal judgments, rankings, and matchings [14].

Consumers are forced into explicit trade-offs in preference elicitation methods like pairwise comparisons or rankings. In such explicit trade-off decisions, consumers are made aware of attribute dependencies at each step of the process. When attribute levels become unavailable, consumers can either change the configuration to make a broader range of attribute levels for the other attributes available, or they can accept the unavailability of some attribute levels and choose within the reduced ranges. They cannot continue the recommendation process before a specific attribute conflict is solved [7]. Implicit trade-offs are required in absolute measurement methods. In implicit trade-off decisions, consumers are not made aware of attribute interdependencies: consumers merely indicate their preferred attribute levels or the importance of attributes. If attribute conflicts arise, the RA interface does not point them out and does not ask to resolve the conflicts. Informed consumers may detect low accuracy in the recommendation output as a result [15].

So far, only few studies have examined the role of trade-offs in RAs. Although explicit consideration of trade-offs apparently improves decision accuracy [8], consumers often attempt to avoid considering trade-offs because they do not like the uncomfortable issue of having to accept (perceived) losses in certain attributes [16–18]. Another drawback of explicit trade-offs is the fact that processing trade-off information is cognitively challenging; depending on the number of attributes, levels and products to be compared, it can lead to information overload. Implicit trade-offs reduce decision stress [14] but choices based on explicit trade-offs are perceived to be more consistent and reliable [15; 19]. While Xu et al. [5] have found consumers to prefer a medium level of trade-off exposure, compared to high and low levels, their results were not confirmed in the context of different preference elicitation methods. We aim to improve the understanding of trade-off exposure by explicitly distinguishing low, medium and high trade-off exposure across different preference elicitation methods.

3 Research Model and Hypothesis Development

We investigate how preference elicitation methods with different levels of trade-off exposure affect consumers' satisfaction with an RA. To study the effect of different levels of trade-off exposure on users' satisfaction with an RA, we propose a research model based on cognitive and affective perceptions of an RA. The choice of latent variables and causal mechanisms is based on previous work, incorporating existing theory on the mechanisms driving users' perception [5; 7]. A summary of the research model is presented in Figure 1.

Trade-off exposure is an environmental cue that affects consumers' cognitive and affective reactions [20]. It becomes apparent to users as a visual cue (aside other cues, e.g. colors or layout), being part of the preference elicitation method implemented in a RA. Cognitively, providing information on unavailable options directly facilitates attaining a shopping goal [2], while the interactive nature of its presentation (e.g. moving sliders or crossed out options) can trigger affective reactions [21].

Perceived control and perceived enjoyment reflect the cognitive and affective reactions to the exposed trade-offs. Perceived control reflects the perception of a consumer's impact on an activity or given condition [22]. In the context of an RA, perceived control expresses to which extent consumers feel they can use the RA to accomplish their intended goal of finding a suitable product. Perceived enjoyment is an affective measure of the consumer's perception whether the interaction with the system is interesting and fun or not [23–25].

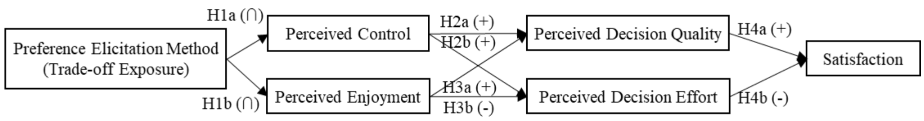


Figure 1. Research Model

Users' final reaction to using the RA is their satisfaction with the recommendation. Prior research has shown that this ultimate response is preceded by perceived decision quality and perceived decision effort [4; 26] – known as the effort-accuracy-framework [27]. Its central idea is that consumers make some compromise between a most accurate decision and the desire to minimize cognitive effort involved. Overall, this research model is based on prior theoretical justification to investigate the role of trade-off exposure as it incorporates both cognitive and affective reactions as well as a rationale for understanding trade-off exposure's effect on consumer satisfaction.

Since preference elicitation methods with low trade-off exposure do not point out attribute conflicts [15], consumers experience discrepancy between current and intended state, leading to lower perceived control [28; 29]. Similarly, the lack of transparency about attribute conflicts may give rise to unrealistic expectations about the decision space, leading to equally unrealistic inputs and in consequence to empty or very small result sets – and in turn, to lower perceived control about the recommendation process.

Trade-offs generally decrease perceived control because they require that "losses" in product attributes, i.e. divergence between current and intended state, are accepted [7; 10]. High trade-off exposure forces the consumer to accept a loss in at least one attribute [16–18] since the consumer cannot continue the recommendation process before a specific attribute conflict is solved. In this case, too, the consumer experiences a discrepancy between current and intended state, resulting in lower perceived control.

A medium level of trade-off exposure makes explicit trade-offs not compulsory but possible. Further, the consumer is informed about changes in attribute level availability after explicitly making a trade-off decision. By giving consumers the choice to only select preferred levels for the attributes of their choice, consumers can apply mitigation strategies. In other words, they can avoid making trade-off decisions between attributes they have little knowledge about and focus on familiar attributes instead. Losses in unfamiliar attributes may be felt less acutely, and the consumer may form more realistic expectations of the decision space. In summary, we expect that:

H1a: Perceived control will follow an inverted U-shape as the level of trade-off exposure increases.

In the context of RAs, the pleasure of using an attractive technology-based tool, as well as the prospect of finding the ideal product, cause enjoyment [30; 31]. An important driver of consumer's perceived enjoyment is the flow state [24]. This concept, central to flow theory, defines a state in which a person is completely focused on an activity, being rewarded just by doing the activity itself, independent of its end result or extrinsic motivation [32].

With respect to enjoyment, low trade-off exposure may leave consumers in a state of underutilizing their knowledge or decision skills. However, "challenges at a level appropriate to one's capacities" are a main antecedent of experiencing a state of flow [32], which drives perceived enjoyment. In contrast to low trade-off exposure, high trade-off exposure may overmatch consumer's cognitive capacity or knowledge, which also prevents flow [32]. RAs with medium trade-off exposure allow consumers to be more active and involved in the process. In addition, consumers can act with a higher degree of self-determination [33] in selecting attributes and proceed in accordance with their capacities. This creates a challenge that matches consumers' existing skills and thereby drives flow [32].

H1b: Perceived enjoyment will follow an inverted U-shape as the level of trade-off exposure increases.

When people perceive high control, they believe that their actions have a higher likelihood of achieving their intended state [28; 29]. In the context of RAs, this translates to the belief that suitable products will be found and suggested by the RA [28]. Consumers who perceive higher control in using the system may expect a higher likelihood of getting a suitable recommendation, hence a higher level of decision quality [34]. When consumers are in full control of the RA, they are able to express the intended state more precisely. They will expect the RA to provide a correspondingly high-quality recommendation [35]. Following these arguments, we propose:

H2a: Higher perceived control leads to increased perceived decision quality.

Higher effort may translate into longer processing times, the need for additional information, or a higher level of energy input (physically and mentally) [36]. At low levels of control, consumers need to find workarounds to bridge the gap between current and intended state, which leads to additional effort [37]. If an RA provides, for instance, insufficient filtering options and consumers are forced to manually search and compare alternatives, they must expend additional effort in terms of both time and cognitive resources. Similarly, the number of iterations required to obtain a satisfactory recommendation affects perceived effort. The more iterations are needed, the more likely it is that consumers must come up with a workaround to produce a state that is different from last iteration and closer to the intended state. This increases cognitive demand since the consumers have to think simultaneously about their actual preferences for the decision task and how to avoid ending up in the same (unsatisfactory) result state as before. Taken together, we suggest that:

H2b: Higher perceived control leads to higher perceived decision effort.

Perceived enjoyment positively affects consumer's future usage intentions [24] as well as their loyalty intentions [38]. In the context of information systems, higher levels of perceived enjoyment are associated with improvements in attitudes and satisfaction with system interfaces [21; 39; 40]. Users who experience higher levels of perceived enjoyment have been found to be more actively involved with the information they process [40]. In terms of a decision task a more active, careful selection of attributes increases the likelihood of ending up with a higher quality product alternative, i.e. higher perceived decision quality. In contrast, not enjoying the decision task leads to being less focused and less careful about the attribute selection and may in turn reduce decision quality perception. Therefore, we expect that:

H3a: Higher perceived enjoyment leads to increased perceived decision quality.

Effort and enjoyment are closely related [41]. For instance, developing software is an effortful task, and software developers are compensated for this effort with wages. However, an astonishingly high number of open source developers are willing to waive such compensation and contribute for free to the open source community [41], motivated by the fun element (i.e. enjoyment) of the task [41]. We expect this effect is observable in decision makers, i.e. they will perceive less effort during a decision, if they enjoy the process. Such perception is related to a flow state where users feel active and creative in contrast to a task experienced as tedious or laborious [42]. Higher perceived enjoyment can also lead consumers to underestimate the difficulty of a task [43]. Consequently, we suggest that:

H3b: Higher perceived enjoyment leads to lower perceived decision effort.

The effort-accuracy framework suggests two main determinants for a positive perception of decision outcome: to increase decision quality and to decrease decision effort [44–46]. This effort-accuracy framework has been widely applied in the research on RAs, for instance to shed light on consumers' intention to reuse RAs [5; 7]. We draw on the same rationale, using satisfaction as a measurement for perception of the decision outcome. Satisfaction has been found to have cognitive [47–49] and affective [50] antecedents [51], which are reflected in perceived decision quality (cognitively) and perceived decision effort (affectively). Harnessing the effort-accuracy framework, we suggest:

H4a: Perceived decision quality positively influences satisfaction.

H4b: Perceived decision effort negatively influences satisfaction.

4 Research Methodology

We conducted a between-subject laboratory experiment to test our research model. The between-subject design was chosen to avoid learning effects and cognitive exhaustion. Participants were randomly assigned to one of three treatments, which were operationalized by different preference elicitation methods with varying levels of trade-off exposure (i.e., low, medium and high). Participants received instructions to use an RA to find the most suitable digital camera they would buy for personal use. We used digital cameras since they are sufficiently complex so that consumers may wish to use an RA [52]. The cameras were composed of 6 attributes with 7 levels each. The product database was identical across treatments. After reading the instructions, participants' attribute preferences were elicited by having them rate and rank 7 random cameras. Subsequently, participants were provided with the respective RA and asked to make a camera purchase decision.

The preference elicitation methods implemented by the RAs were taken from literature, each corresponding in its characteristics to the specifications outlined in section 2 for the respective level of trade-off exposure. Figure 2 illustrates all three preference elicitation methods. In the low trade-off exposure condition, participants were described a situation in which they are stuck between alternatives and asked to assign 100 points to the most important attribute [53]. Participants were then asked to continue with the second most important attribute and assign a lower number of points than to the previous one. This process continued until all attributes had been assigned points.

Trade-off exposure is low because no explicit trade-off between two attributes is supported, and no feedback is given on how the order or the points assigned affect the decision space.

In the high trade-off exposure condition, participants were asked to compare two product alternatives that differed in only two attributes. They were then told to adjust the level of one attribute until both alternatives are equally preferred. This principle is used widely and in many variations [6; 54]. Figure 2 presents the different preference elicitation methods and illustrates high trade-off exposure on the example of zoom and Megapixels (MP).

In the medium trade-off exposure condition, participants were presented with a configuration interface [9] and asked to change attribute levels according to their preferences. The RA determined the price interval for each new configuration and disabled unavailable levels for other attributes. This approach allows for mitigation strategies but enables the participant to perform explicit trade-offs, if required, corresponding to the requirements for a medium level of trade-off exposure.

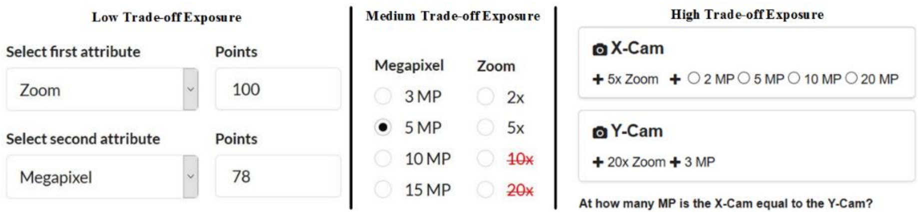


Figure 2. Treatment interfaces

In the questionnaire following the treatment, we collected six variables regarding RA perceptions measured on a 7-point Likert scale. We elicited perceived enjoyment [41; 55], perceived control [55], perceived decision effort [55], perceived decision quality [5; 7; 19] and satisfaction [51]. In addition, participants gave their age, gender and prior experience with RAs and a manipulation check was conducted.

116 students from the University of Passau took part in the laboratory experiment and were assigned randomly to the three treatments (40 participants to low trade-off exposure, 34 to medium trade-off exposure and 42 to high trade-off exposure).

Each participant received a payoff of 10 euros. 36% of the participants were female. Participants' age ranged from 18 to 46 ($\mu = 23$, $\sigma = 5.00$). Familiarity with RAs was average ($\mu = 3.20$, $\sigma = 0.95$). None of the control variables had a significant effect on satisfaction (linear regression results: $p_{age} > 0.1$, $p_{gender} > 0.1$, $p_{experience} > 0.1$).

5 Analysis and Results

We use a two-step analysis approach as previously applied in related studies [5; 7]. First, we examine the influence of trade-off exposure on perceived control and perceived enjoyment (H1a and H1b; Kruskal-Wallis-test and Mann-Whitney-U-tests). Second, we examine the effects of perceived control on effort and quality and how the latter affect satisfaction. (H2-H4; PLS path modelling). We chose Mann-Whitney-U-

tests because Kolmogoroff-Smirnoff tests indicated that our data was not normally distributed ($p_{control} < 0.001$, $p_{enjoyment} < 0.001$) [56].
We first compared perceived control across the three levels of trade-off exposure using a Kruskal-Wallis test, indicating that values of different groups have significantly different distributions ($X^2 = 33.35, p < 0.001$). For perceived enjoyment, we found significant group differences as well ($X^2 = 11.08, p < 0.01$). To shed light on how the values are different across groups, we used pairwise, one-sided Mann-Whitney-U tests between low and medium trade-off exposure and between medium and high trade-off exposure, respectively. Perceived control had its highest mean of 4.87 for medium trade-off exposure: it was significantly higher than in a low trade-off exposure condition ($4.87 > 4.00, p < 0.001$) as well as in high trade-off exposure ($4.87 > 3.52, p < 0.001$), supporting H1a. For perceived enjoyment, the mean in the medium trade-off exposure group was also significantly higher than in low ($4.32 > 3.76,\ p < 0.01$) and high ($4.32 > 3.59,\ p < 0.01$) trade-off exposure groups. This supports the inverted-u shaped relationship proposed in H1b, which is also illustrated in Figure 3.

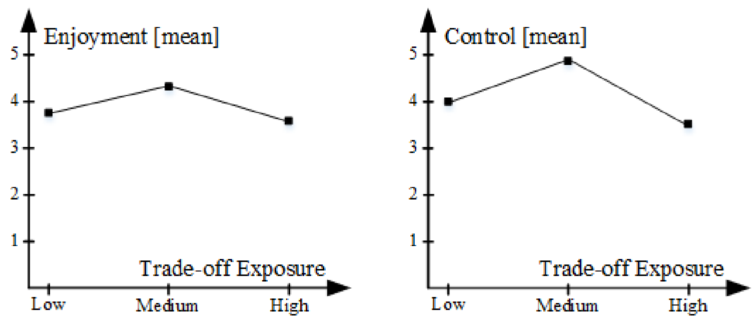


Figure 3. Visualization of U-shaped relationships

Turning to H2-H4, we used a PLS path model to investigate the relationships leading to consumers' satisfaction perception. We analyzed the PLS path model following the two-step process of outer and inner model assessment [57]. This approach is used to ensure reliable results as the analysis of the inner models' paths relies on the reliability and validity of the outer models' constructs [58]. Following Ringle et al. [59]'s call for thorough PLS reporting, we would like to clarify that the analyses were carried out with R 3.1.3, using the plspm library with standard settings.
First, we checked for individual item reliability. Each item should show a substantial correlation with its construct. For reflective items, as solely present in our model, this can be assessed by item loading, i.e. its variance explained by the construct. Loadings should not be lower than 0.7 [60], which is fulfilled in our model. Second, internal consistency was assessed based on item intercorrelations, measured by Cronbach's alpha [61] and composite reliability [62]. Values exceed the recommended threshold of 0.7 for all constructs [63]. Taken together, these results indicate that the measurement model is reliable.
We then assessed convergent validity, i.e. whether all items in a construct's block unidimensionally represent their construct [58]. As recommended by Fornell and Larcker [63] we calculated the average variance extracted (AVE) for every latent

variable. Table 1 shows that AVE is greater than the recommended threshold of 0.5 for every construct [64].
Finally, discriminant validity was examined to confirm that different latent variables actually exhibit significant difference [58]. We checked the Fornell-Larcker criterion [63], stating that the AVE of each latent variable should be greater than its squared correlation with any other variable (which is equal to the AVE's square root being greater than the variables' correlations, which we used for simpler reporting). This is the case, as the values in Table 1 confirm. We also checked the Heterotrait-Monotrait Matrix (HTMT). This criterion has shown superior performance in detecting discriminant validity in a Monte Carlo simulation study compared to cross-loadings analysis and the Fornell-Larcker criterion [65]. The computed HTMT values for our model can be found in Table 1. All values are below 0.85, which satisfies $H_{0.85}$, the HTMT criterion with highest specificity [65]. We conclude that overall measurement model validity has been established.

Table 1. Structural model discriminant validity measures

	Enjoyment	Control	Quality	Effort	Satisfaction
Enjoyment	0.893	0.440	0.373	0.433	0.569
Control	0.397	0.949	0.392	0.421	0.727
Quality	0.346	0.354	0.921	0.320	0.751
Effort	-0.399	-0.383	-0.290	0.876	0.570
Satisfaction	0.509	0.631	0.690	-0.518	0.844
Latent variable correlations		$\sqrt{\text{AVE}}$		HTMT matrix	

Analyzing the structural model next, we focus on perceived control first: We found a significant positive influence on perceived decision quality ($\beta = 0.257$; $p < 0.01$), supporting H2a. Moreover, there was a positive influence of perceived control on perceived decision effort ($\beta = 0.267$; $p < 0.01$) supporting H2b. Higher levels of perceived enjoyment were found to lead to an increase in perceived decision quality ($\beta = 0.245$; $p < 0.01$), while perceived decision effort was reduced ($\beta = -0.293$; $p < 0.01$). Hence, hypotheses H3a and H3b are supported. Turning to the effort-accuracy framework, perceived decision quality had a significant positive effect on satisfaction ($\beta = 0.590$; $p < 0.001$), supporting H4a. Finally, higher perceived decision effort had a significant negative effect on satisfaction ($\beta = -0.347$; $p < 0.001$), which supports H4b.

6 Discussion

Prior research has often made only a dichotomous distinction regarding the level of trade-off exposure, comparing high and low levels while neglecting a medium option. While Xu et al. [5] have showed the possibility for such a medium level, their study did focus on a single preference elicitation method. Our study makes a contribution to the understanding of the relevance of medium trade-off exposure as we investigated how preference elicitation methods with different levels of trade-off exposure affect consumers' satisfaction with an RA. The results support our supposition that medium-level trade-off exposure generates the highest levels of perceived control and perceived enjoyment, compared to low or high trade-off exposure, and that perceived control and perceived enjoyment follow an inverted U-shaped curve as the level of trade-off exposure increases. These findings conform the results of Xu et al. [5], however, in a different research setting and across different preference elicitation methods. Note, however, that preference elicitation methods with low and high trade-off exposure also achieved good levels of perceived enjoyment and control.

The results of our study are consistent with prior research on consumers' perceptions towards RAs, showing that stimuli like trade-off exposure affect cognitive and affective reactions, and those in turn affect the response of the customer. Specifically, our findings show that perceived decision quality was positively influenced by perceived control and perceived enjoyment, while perceived decision effort was affected negatively. In line with the effort-accuracy model, perceived decision quality had a positive effect and perceived decision effort a negative effect on consumer satisfaction. Overall, our results show that a preference elicitation method with a medium level of trade-off exposure creates highest consumer satisfaction with a RA.

6.1 Implications for theory and practice

On a theoretical level, our study contributes to a better understanding of the role of trade-offs in RAs. Specifically, this is the first study to investigate the effect of different levels of trade-off exposure as realized by different preference elicitation methods. The few studies that simultaneously investigated different levels of trade-off exposure mostly focused on high and low levels, neglecting a medium level option. Thus until now, researchers used a linear model for their studies and did not investigate a non-linear coherence of trade-off exposure and the acceptance of it. This study contributes to this knowledge gap by differentiating between high, medium and low trade-off exposure and points out that a medium level has the highest effect on affective and cognitive perceptions.

Our study also contributes to the understanding of the processes that affect satisfaction with an RA. Our research model combines perceived enjoyment and perceived control which each have been part of separate models on consumers' RA perception [5; 7]. Also, in contrast to these models, we have used satisfaction as a more proximate response variable as it is known to be the main determinant of reuse intention [66]. Our model helps to understand consumers' cognitive and affective processes during RA usage and can be used to investigate consumer responses to other RA cues.

Our study also provides practical implications for RA developers. Our results suggest that it is better to avoid using only implicit trade-offs but that forced trade-offs may overburden consumers. Rather, RA developers ought to devise ways of implementing medium trade-off exposure, for instance by using configuration-based preference elicitation methods [5; 9]. The increased influence of consumer actions on the elicitation process in combination with the interactive feedback design can help to leverage the positive effects of a medium level of trade-off exposure.

6.2 Limitations, Future Research

This research is subject to several limitations, to be addressed in future research. First, our experiment was conducted in a laboratory setting. Replicating the study in a field setting would help to establish external validity of our findings. Second, other consumer groups, e.g. older or less technology-savvy consumers, may show different reactions than the student sample used in this study. Further studies with other demographical groups and with other products are required for establishing generalizability of our results. Third, we have focused on trade-off exposure; however, there are multiple other aspects to differing preference elicitation methods. Further studies may try to isolate trade-off exposure or widen the aspects under investigation to provide a more fine-grained understanding of preference elicitation and trade-off exposure.

6.3 Conclusion

This study contributes to the understanding of trade-off exposure in RAs. We compare low, medium and high levels of trade-off exposure operationalized in three distinct preference elicitation methods. The results show that both perceived enjoyment and control follow an inverted u-shape when trade-off exposure increases. While low and high levels also yield good performances, a medium level of trade-off exposure leads to significantly highest results. We further presented a research model that shows how trade-off exposure influences consumers' satisfaction with a RA. While perceived enjoyment and perceived control are directly affected by trade-off exposure, they in turn affect perceived decision effort and perceived decision quality as central determinants of overall satisfaction.

References

1. Broniarczyk, S. M., Griffin, J. G.: Decision Difficulty in the Age of Consumer Empowerment. Journal of Consumer Psychology, 24(4), 608–625 (2014)
2. Parboteeah, D. V., Valacich, J. S., Wells, J. D.: The influence of website characteristics on a consumer's urge to buy impulsively. Information Systems Research, 20(1), 60–78 (2009)
3. Goldstein, W. M., Barlas, S., Beattie, J.: Talk about Tradeoffs. Judgements of Relative Importance and Contingent Decision Behavior. In: Weber, EU, Baron, J, Loomes, G (Hrsg), *Conflict and Tradeoffs in Decision Making.* Cambridge University Press, New York, NY, US (2001)
4. Häubl, G., Murray, K. B.: Preference Construction and Persistence in Digital Marketplaces. The Role of Electronic Recommendation Agents. Journal of Consumer Psychology, 13(1/2), 75–91 (2003)
5. Xu, J., Benbasat, I., Cenfetelli, R. T.: The Nature and Consequences of Trade-off Transparency in the Context of Recommendation Agents. MIS Quarterly, 38(2), 379–406 (2014)
6. Pöyhönen, M., Hämäläinen, R. P.: On the convergence of multiattribute weighting methods. European Journal of Operational Research, 129(3), 569–585 (2001)
7. Lee, Y. E., Benbasat, I.: Research Note—The Influence of Trade-off Difficulty Caused by Preference Elicitation Methods on User Acceptance of Recommendation Agents Across Loss and Gain Conditions. Information Systems Research, 22(4), 867–884 (2011)

8. Frisch, D., Clemen, R. T.: Beyond Expected Utility. Rethinking Behavioral Decision Research. Psychological Bulletin, 116(1), 46–54 (1994)
9. Scholz, M., Dorner, V., Schryen, G., Benlian, A.: A configuration-based recommender system for supporting e-commerce decisions. European Journal of Operational Research, 259(1), 205–215 (2017)
10. Bagozzi, R. P.: Principle of Marketing Management. SRA, Chicago (1986)
11. Xiao, B., Benbasat, I.: E-Commerce Product Recommendation Agents. Use, Characteristics, and Impact. MIS Quarterly, 31(1), 137–209 (2007)
12. Yue, S., Larson, M., Hanjalic, A.: Collaborative filtering beyond the user-item matrix. A survey of the state of the art and future challenges. ACM Computing Surveys, 47(1), 1–45 (2014)
13. Huang, S.-L.: Designing utility-based recommender systems for e-commerce. Evaluation of preference-elicitation methods. Electronic Commerce Research and Applications, 10(4), 398–407 (2011)
14. Aloysius, J. A., Davis, F. D., Wilson, D. D., Ross Taylor, A., Kottemann, J. E.: Decision Aiding. User Acceptance of Multi-criteria Decision Support Systems: The Impact of Preference Elicitation Techniques. European Journal of Operational Research, 169, 273–285 (2006)
15. Fasolo, B., McClelland, G. H., Lange, K. A.: The Effect of Site Design and Interattribute Correlations on Interactive Web-Based Decisions. In: Haugtvedt, CP, Machleit, KA, Yalch, RF (Hrsg), *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (2005)
16. Hogarth, R. M.: Judgement and Choice. The Psychology of Decision. John Wiley & Sons (1987)
17. Tetlock, P. E.: The Impact of Accountability on Judgment and Choice. Toward a Social Contingency Model. Advances in Experimental Social Psychology, 25(3), 331–376 (1992)
18. Tversky, A., Shafir, E.: Choice Under Conflict. The Dynamics of Deferred Decision. Psychological Science, 3(6), 358–361 (1992)
19. Widing, R. E., Talarzyk, W. W.: Electronic Information Systems for Consumers. An Evaluation of Computer-Assisted Formats in Multiple Decision Environments. Journal of Marketing Research, 30(2), 125–141 (1993)
20. Eroglu, S. A., Machleit, K. A., Davis, L. M.: Atmospheric Qualities of Online Retailing. A Conceptual Model and Implications. Journal of Business Research, 54(2), 177–184 (2001)
21. Jiang, Z., Benbasat, I.: nvestigating the Influence of Interactivity and Vividness on Online Product Presentations. Information Systems Research, 18(4), 454–470 (2007)
22. Frese, M.: A Theory of Control and Complexity. Implications for Software Design and Integration of Computer Systems into the Work Place. In: Michael, F, Eberhard, U, Wolfgang, D (Hrsg), *Psychological Issues of Human-computer Interaction in the Work Place.* North-Holland Publishing Co. (1987)
23. Csikszentmihalyi, M.: Beyond Boredom and Anxiety. Jossey-Bass, San Francisco (1977)
24. Koufaris, M.: Applying the technology acceptance model and flow theory to online consumer behavior. Information Systems Research, 13(2), 205–223 (2002)
25. Benlian, A.: Web Personalization Cues and Their Differential Effects on User Assessments of Website Value. Journal of Management Information Systems, 32(1), 225–260 (2015)
26. Hostler, E. R., Yoon, V. Y., Guimaraes, T.: Assessing the impact of internet agent on end users' performance. Decision Support Systems, 41(1), 313–323 (2005)
27. Payne, J. W., Bettman, J. R., Johnson, E. J.: The Adaptive Decision Maker. Cambridge University Press (1993)

28. Schmitz, B., Skinner, E.: Perceived control, effort, and academic performance. Interindividual, intraindividual, and multivariate time-series analyses. Journal of Personality and Social Psychology, 64(6), 1010–1028 (1993)
29. Skinner, E., Greene, T.: Perceived control. Engagement, coping, and development. 21st century education: A reference handbook, 1, 121–130 (2008)
30. Dabholkar, P. A.: Consumer evaluations of new technology-based self-service options. An investigation of alternative models of service quality. International Journal of Research in Marketing, 13(1), 29–51 (1996)
31. Dabholkar, P. A., Bagozzi, R. P.: An Attitudinal Model of Technology-Based Self-Service. Moderating Effects of Consumer Traits and Situational Factors. Journal of the Academy of Marketing Science, 30(3), 184–201 (2002)
32. Nakamura, J., Csikszentmihalyi, M.: The concept of flow. In: , *Flow and the foundations of positive psychology.* Springer (2014)
33. Deci, E. L., Ryan, R. M.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American psychologist, 55(1), 68 (2000)
34. Jarupathirun, S., Zahedi, F.: Exploring the Influence of Perceptual Factors in the Success of Web-based Spatial DSS. Decision Support Systems, 43(3), 933–951 (2007)
35. Kramer, T.: The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. Journal of Marketing Research, 44(2), 224–233 (2007)
36. Smith, V. L., Walker, J. M.: Monetary rewards and decision cost in experimental economics. Economic Inquiry, 31(2), 245–261 (1993)
37. Norman, D.: Emotion & design. attractive things work better. interactions, 9(4), 36–42 (2002)
38. Cyr, D., Head, M., Ivanov, A.: Perceived interactivity leading to e-loyalty. Development of a model for cognitive–affective user responses. International Journal of Human-Computer Studies, 67(10), 850–869 (2009)
39. Morrison, J., Vogel, D.: The Impacts of Presentation Visuals on Persuasion. Information & Management, 33(3), 125–135 (1998)
40. Griffith, D. A., Krampf, R. F., Palmer, J. W.: The Role of Interface in Electronic Commerce. Consumer Involvement with Print versus On-line Catalogs. International Journal of Electronic Commerce, 5(4), 135–153 (2001)
41. Franke, N., Schreier, M.: Why Customers Value Self- Designed Products. The Importance of Process Effort and Enjoyment. Journal of Product Innovation Management, 27(7), 1020–1031 (2010)
42. Csikszentmihalyi, M., LeFevre, J.: Optimal Experience in Work and Leisure. Journal of Personality and Social Psychology, 56(5), 815 (1989)
43. Agarwal, R., Karahanna, E.: Time flies when you're having fun. Cognitive absorption and beliefs about information technology usage. MIS Quarterly, 24(4), 665–694 (2000)
44. Todd, P., Benbasat, I.: The Use of Information in Decision Making. An Experimental Investigation of the Impact of Computer-Based Decision Aids. MIS Quarterly, 16(3), 373–393 (1992)
45. Häubl, G., Trifts, V.: Consumer Decision Making in Online Shopping Environments. The Effects of Interactive Decision Aids. Marketing Science, 19(1), 4–21 (2000)
46. Wang, W., Benbasat, I.: Interactive Decision Aids for Consumer Decision Making in E-commerce. The Influence of Perceived Strategy Restrictiveness. MIS Quarterly, 33(2), 293–320 (2009)
47. Oliver, R. L.: A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. Journal of Marketing Research, 17(4), 460–469 (1980)

48. LaBarbera, P. A., Mazursky, D.: A Longitudinal Assessment of Consumer Satisfaction/Dissatisfaction. The Dynamic Aspect of the Cognitive Process. Journal of Marketing Research, 20(4), 393–404 (1983)
49. Oliver, R. L., DeSarbo, W. S.: Response Determinants in Satisfaction Judgments. Journal of Consumer Research, 14(4), 495–507 (1988)
50. Westbrook, R. A., Oliver, R. L.: The dimensionality of consumption emotion patterns and consumer satisfaction. Journal of Consumer Research, 18(1), 84–91 (1991)
51. Homburg, C., Koschate, N., Hoyer, W. D.: The Role of Cognition and Affect in the Formation of Customer Satisfaction. A Dynamic Perspective. Journal of Marketing, 70(3), 21–31 (2006)
52. Wang, W., Benbasat, I.: Trust in and Adoption of Online Recommendation Agents. Journal of the Association of Information Systems, 6(3), 72–101 (2005)
53. Winterfeldt, D. von, Edwards, W.: Decision analysis and behavioral research. Cambridge University Press Cambridge (1986)
54. Keeney, R. L., Raiffa, H.: Decisions with Multiple Objectives. Preferences and Value Trade-Offs. Wiley, New York (1976)
55. Dellaert, B. G. C., Dabholkar, P. A.: Increasing the Attractiveness of Mass Customization. The Role of Complementary On-line Services and Range of Options. International Journal of Electronic Commerce, 13(3), 43–70 (2009)
56. Conover, W. J.: Practical Nonparametric Statistics. 3. Auflage. Wileys (1980)
57. Chin, W. W.: The Partial Least Squares Approach to Structural Equation Modeling. Modern Methods for Business Research, 295(2), 295–336 (1998)
58. Henseler, J., Ringle, C. M., Sinkovics, R. R.: The Use of Partial Least Squares Path Modeling in International Marketing. Advances in International Marketing (AIM), 20, 277–320 (2009)
59. Ringle, C. M., Sarstedt, M., Straub, D. W.: A Critical Look at the Use of PLS-SEM in MIS Quarterly. MIS Quarterly, 36(1), 4–8 (2012)
60. Barclay, D., Higgins, C., Thompson, R.: The Partial Least Squares (PLS) Approach to Causal Modeling. Personal Computer Adoption and Use as an Illustration. Technology Studies, 2(2), 285–309 (1995)
61. Cronbach, L. J.: Coefficient Alpha and The Internal Structure of Tests. Psychometrika, 16(3), 297–334 (1951)
62. Werts, C. E., Linn, R. L., Jöreskog, K. G.: Intraclass Reliability Estimates. Testing Structural Assumptions. Educational and Psychological Measurement, 34(1), 25–33 (1974)
63. Fornell, C., Larcker, D. F.: Structural Equation Models with Unobservable Variables and Measurement Error. Algebra and Statistics. Journal of Marketing Research, 18(3), 382–388 (1981)
64. Götz, O., Liehr-Gobbers, K., Krafft, M.: Evaluation of Structural Equation Models Using the Partial Least Squares (PLS) Approach. In: Esposito Vinzi, V., Chin, W.W., Henseler, J., Wang, H. (Hrsg), *Handbook of Partial Least Squares.* Springer (2010)
65. Henseler, J., Ringle, C. M., Sarstedt, M.: A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling. Journal of the Academy of Marketing Science, 43(1), 1–21 (2015)
66. DeLone, W. H., McLean, E. R.: The DeLone and McLean Model of Information Systems Success. A Ten-Year Update. Journal of Management Information Systems, 19(4), 9–30 (2003)

Theory-Based Affordances of Utilitarian, Hedonic and Dual-Purposed Technologies: A Literature Review

Dorothee Wittek[1], Manuel Wiesche[2], Klaus Goffart[1], and Helmut Krcmar[2]

[1] BMW Group, Munich, Germany
{dorothee.wittek, klaus.goffart}@bmw.de
[2] Technical University of Munich, Department of Information Systems, Munich, Germany
{manuel.wiesche, krcmar}@in.tum.de

Abstract. The key to an information system's (IS) success is its value experienced by the user. A promising approach to enhance user value is to design for the users' experiential desires. For example, fulfilled experiential desires enhance the users' satisfaction and loyalty. Despite these benefits, few design principles exist for developing IS according to the users' experiential desires. Therefore, the aim of this literature review is to aggregate the current state of knowledge concerning the different theoretical perspectives on utilitarian, hedonic and dual-purposed IS. We build a framework that illustrates the relationship between different theoretical perspectives on IS affordances (i.e., motivational, cognitive, affective, and social) and different technology types (i.e., utilitarian, hedonic, dual-purposed). The presented framework offers a starting point for the development of theory-based design principles for experiential affordances of IS. We conclude with a summary of opportunities for future research to extend our knowledge of experientially fulfilling IS.

Keywords: Motivational affordances, dual-purposed systems, hedonic systems, utilitarian systems.

1 Introduction

The key to an information system's success is its value experienced by the user [1]. Current research mainly distinguishes instrumental and experiential values [2]. Instrumental values contain pragmatic or utilitarian product qualities and are linked to instrumental outcomes such as enhanced productivity or reduced expenditures. Experiential values contain hedonic product qualities and are linked to experiential outcomes such as meaningfulness, engagement, positive emotions or perceived enjoyment [2]. According to their main value and outcome, different technologies can be classified as three different technology types, namely (1) predominantly utilitarian, (2) predominantly hedonic and (3) hybrid or dual-purposed [3]. Utilitarian technologies are mainly used at the workplace or in productivity-oriented contexts of use and provide instrumental value. In contrast, predominantly hedonic technologies are mainly used during leisure time or at home and provide fun and entertainment [4, 5]. Dual-purpose

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

technologies are, however, a hybrid of utilitarian and hedonic technologies [3]. A prominent example of dual-purpose systems are social networking systems which are commonly used for either leisure or work purposes (e.g., Xing).

Since the last decade, especially hedonic and dual-purposed technologies represent an emerging type of IS. Thereby, experiential outcomes are receiving growing attention in research and practice [2, 6]. For example, gamified design elements are used as means of providing experiential value and are already integrated in many products such as internet portals and cars. Here, drivers are for example motivated to improve their fuel economic driving by gamified elements [7, 8]. The interest in the experiential perspective on technology use results from its promising benefits such as enhancing important facets of technology acceptance like for example the users' satisfaction [9], word-of-mouth [10] and behavioral intention [11]. Additionally, experiential values enhance instrumental outcomes such as perceived ease of use which will in turn again enhance outcomes of technology acceptance [9]. Moreover, experiential values enable behavioral change such as reduced energy consumption [12].

In contrast to the expected growth of experiential value [e.g., 6], attempts to design for experiential outcomes fail quite often [e.g., 13]. A prominent example from the organizational context are public leaderboards [14]. For example, managers at DisneyLand tried to motivate their employees with public leaderboards of the most efficient employees. Instead of being motivated and experiencing a gamified competition, the employees mostly felt very uncomfortable with this idea of gamification. Moreover, market analysts discussed gamification trends and strategies and concluded that "80% of current gamified applications will fail to meet business objectives primarily due to poor design" [15]. Furthermore, the MISQ recently published a call for research that stated that "few research and design guidelines exist regarding gamified information systems" and called "to investigate the design and use of gamified information systems from a variety of disciplinary perspectives and theories, including behavioral economics, psychology, social psychology, information systems" [2]. Therefore, we argue that the high failure rate of attempts to design for experiential outcomes is due to the problem that few design principles exist for developing IS according to the users' experiential desires. According to [2], we define design principles as high-level design rules and formulas that should be derived from grounded theory and can support product developers through the whole development process. For example, a design principle in the field of gamification might suggest that gamified IS should include different badges depending on the different user styles and stages. For the creation, application, and evaluation of theory-based design principles that can speak directly to the users' motivation, cognition, affect and social behavior, it is necessary to use suitable theoretical foundations. Therefore, we need to analyze basic research, theories and models from a variety of disciplines like information systems (IS), behavioral economics, human-computer interaction and psychology that can be used to derive experiential affordances. Therefore, the aim of this review is to aggregate the current state of knowledge concerning the different theoretical perspectives on utilitarian, hedonic and dual-purposed IS. We build a framework that illustrates the relationship between different theoretical perspectives on IS affordances (i.e., motivational, cognitive, affective, and social) and technology types (i.e., utilitarian, hedonic, dual-purposed). The presented framework offers a starting point for the development of theory-based design principles for experiential affordances of IS.

Our review is structured as follows: first, we describe the design of our literature review. This includes the selection of appropriate databases, journals and conference proceedings and the coding of the identified relevant papers according to their main theoretical perspective. Second we provide an overview of each theoretical perspective based on our presented concept matrix. Third, we summarize knowledge gaps and opportunities for future research. Finally, our review ends with a conclusion on theoretical and practical implications.

2 Design of the Literature Review

The aim of this paper is to understand and aggregate the current state of knowledge concerning the different theoretical perspectives on utilitarian, hedonic and dual-purposed IS. Thereby, we build a framework that illustrates the relationship between different theoretical perspectives on IS affordances (e.g., motivational theory perspective) and different technology types (i.e., utilitarian, hedonic, dual-purposed). Therefore, we conducted a systematic literature review based on the guidelines of Webster and Watson [16]. We combine research from a variety of disciplines including IS, behavioral economics, human-computer interaction and psychology. An overview of the scope of our review, our search terms and the considered databases can be found in Table 1. In order to decide which papers were relevant for our review, we focused on the following two inclusion criteria: (1) we only included papers that focus on interactive technology because we are interested in designing for outcomes that users derive from the direct interaction with technologies; (2) we only included papers that concentrate on voluntary use of technology because we are interested in the consumer context and not the enterprise software context. Hence, we excluded papers that focused on the organizational context or non-voluntary use of IS in the educational context. In order to identify all relevant papers, we screened the title, abstract and if necessary the whole paper. Finally, including the results of our forward and backward search, 42 papers remained as relevant hits in our review.

Based on Webster and Watson [16] we created a concept matrix to structure our findings. Since our review is meant to provide an overview about the existing theoretical perspectives on utilitarian, hedonic and dual-purposed IS, we structured the relevant papers according to the considered type of technology, namely (1) utilitarian, (2) hedonic, and (3) dual-purposed. Moreover, we tried to find a structure to group different theories into one concept matrix. By filling out our concept matrix, we iteratively refined our columns and finally focused on four main branches of theories, namely (1) motivational, (2) cognitive, (3) affective, and (4) social theoretical perspectives. Motivational theories in the context of technology use [e.g., 17] mainly focus on the interplay of product characteristics and different kinds of human motivations (e.g., intrinsic vs. extrinsic motivation). Cognitive theories mainly focus on the cognitive processing of product characteristics, decision-making processes and product choice scenarios [e.g., 9, 18]. The core element of affective theories is the role of human emotions in the perception, use and evaluation of technologies [e.g., 19].

Finally, social theories mainly concentrate on the influence of social interaction and response patterns on technology use [e.g., 20].

Table 1. Systematic Literature Search Process

Database	*Search Term*	*Search Fields*	*Hits*	*Relevant*
ScienceDirect	("hedonic" OR "experiential" OR "enjoyment") AND ("pragmatic" OR "utilitarian" OR "instrumental")	Title, Abstract and Keywords	145	18
EbscoHost			576	7
ICIS			23	4
ECIS			9	1
			Sum	30
			Forward Backward Search	12
			Total	42

3 Findings

Table 2 illustrates a selection of our concept matrix. In total, 42 papers were clustered according to their main theoretical perspective (i.e., motivational, cognitive, affective, social) and the considered technology type (i.e., utilitarian, hedonic, dual-purposed). Figure 1 illustrates all theories we identified as theoretical basis in the studies that were part of our systematic literature review. In the following paragraphs, we provide an overview of the motivational, cognitive, affective, and social perspective on IS affordances and refer to a selection of the theories listed in Figure 1.

Table 2. Selection of Concept Matrix (Mot. = Motivational, Cog. = Cognitive, Aff. = Affective, Soc. = Social, Util. = Utilitarian, Hed. = Hedonic, D-P = Dual-purposed)

	Theory				*Technology Type*		
Source	Mot.	Cog.	Aff.	Soc.	Util.	Hed.	D-P
[22]					x		
[24]					x		
[27]	x		x	x		x	
[28]		x					x
[29]	x		x		x	x	x
[30]	x				x	x	
[31]		x	x		x	x	
[32]					x	x	
[36]	x	x			x		
[37]						x	
[…]							
Total	20	21	11	7	21	26	23

Theoretical Perspecitves on IS Affordances

Social:

1) Cooperative Principle of Conversation
2) Optimal Stimulation*
3) Persuasive Systems Design Model
4) Social Comparison Theory
5) Social Cognitive Theory*
6) Social Influence & Norms
7) Social Interaction Studies
8) Social Response Theory

Affective:

9) Affect and Emotion Studies
10) Affect Control Theory*
11) Affective Computing
12) Cognitive Absorbtion*
13) Cue-Utilization Theory*
14) Decomposed Theory of Planned Behavior*
15) Emotional Design Paradigm
16) Exploration Theory*
17) Fear Activation Model*
18) Flow Theory
19) Hedonic Theory*
20) Model for Hedonic Information System Acceptance*
21) M-R Framework / Stimulus-Organism-Model*
22) PAD Theory (Pleasure, Arousaal, Dominance)
23) Product Meaning and Choice Mode: Triangle of Product Meaning
24) Theory of Planned Behavior*
25) Theory of Reasoned Action*

Cognitive:

10) Affect Control Theory*
34) Cognitive Evaluation Theory (Self-Perception Theory)
12) Cognitive Absorption*
35) Cognitive Design Principles
36) Cognitive Dissonance Theory
37) Cognitive Fit Theory
13) Cue-Utilization Theory*
14) Decomposed Theory of Planned Behavior*
17) Fear Activation Model*
38) Feature Detection Perception Process
19) Hedonic Theory*
39) Information-Processing Model of Consumer Decision-Making Process and Satisfaction
20) Model for Hedonic Information System Acceptance*
21) M-R Framework / Stimulus-Organism-Model*
40) Process of Elaboration: Semantic (a) and Autobiographic (b) Elaboration
41) Product Choice & Justifiability
42) Product Meaning and Choice Mode: Triangle of Product Meaning
43) Prospect Theory
44) Rational Choice Theory
5) Social Cognitive Theory*
45) Theories on Mental Workload
24) Theory of Planned Behavior*
25) Theory of Reasoned Action*
46) Expectation-Confirmation Model/Paradigm

- Adaptation Level Theory
- Anchoring Theory
- Design-Expectation Fit
- Expectancy Theory
- Expectation Disconfirmation Paradigm
- Expectations-Disconfirmation Theory
- Met-Expectations Theories

Motivational:

16) Exploration Theory*
26) Hedonic-Motivation System Adoption Model
27) Herzberg's Motivation Theory / Herzberg's Two-Factor Theory: Motivators (a) vs. Hygiene Factors (b)
28) Motivational Affordances
29) Unified Theory of Acceptance and Use of Technology (UTAUT)
30) Motivational Theory: Intrinsic (a) vs. Extrinsic Motivation (b)
31) Multimotive Information Systems Continuance Model
2) Optimal Stimulation*
32) Self-Determination-Theory
33) Technology Acceptance Model (TAM)

Figure 1. Theory Mind Map (* = theories that were grouped into more than one category)

3.1 Motivational Perspective on IS Affordances

In total, we found 20 papers that focused on a motivational perspective on IS affordances. Motivational affordances are defined as the "properties that afford user motivation" [38: p. 274] and are seen as a "key requirement for behavior change" [38: p. 271]. Here, especially Herzberg's [39] Motivator-Hygiene-Theory and Deci's [40] distinction between the two fundamental types of intrinsic and extrinsic motivations are frequently mentioned and applied. While motivators are seen as IS characteristics that provide satisfaction if fulfilled, hygiene factors only cause dissatisfaction if not fulfilled. In the context of intrinsic and extrinsic motivation, user acceptance is seen as either driven by benefits derived from an engaging interaction with the system per se (i.e., intrinsic) or by expected benefits of external rewards (i.e., extrinsic).

Two newer developments within the group of motivational theories are for example the Hedonic-Motivation System Adoption Model (HMSAM) [34] and the Multimotive Information Systems Continuance Model (MISC) [35]. Both models are originally based on the distinction of intrinsic and extrinsic motivators [40]. The HMSAM is meant to improve the understanding of the adoption of hedonic-motivation systems and therefore integrates flow-based cognitive absorption as a mediator of perceived ease of use and behavioral intention. The MISC focuses on the users' expectations and disconfirmations as antecedent of behavioral intention.

Two very concrete examples of applying the motivational perspective are presented by Resatsch [41] and Füller [42]. For example, Resatsch [41] focused on motivating applications in the field of ubiquitous computing in the office, retail and ticketing context and formulated and evaluated design guidelines for NFC-based ubiquitous computing applications. Füller [42] concentrated on designing IT-based customer integration methods and created a framework for positive customer integration experience based on the Motivator-Hygiene-Theory [39].

The motivational perspectives also contain cognitive, social and affective components as for example intrinsic motivators are often conceptualized as emotions like fun, enjoyment, playfulness, pleasure, arousal or dominance [5]. Moreover, motivational needs are often conceptualized as psychological (i.e., autonomy, competence, and relatedness) or social needs (i.e., achievement, affiliation and intimacy, and leadership and followership) [38]. Therefore, the perspectives presented in the following sections are closely related to the paramount motivational perspective.

3.2 Cognitive Perspective on IS Affordances

Most of the identified studies referred to a cognitive or a combined cognitive and affective perspective on IS affordances. Within this category, the Theory of Reasoned Action [43] and the Theory of Planned Behavior [44] are the basis of several models of IS affordances [e.g., 21, 23]. Here, the affective and behavioral reactions towards an IS are seen as the result of cognitive processes including attitudes, subjective norms and perceived behavioral control. Cognitive processes can for example include the semantic and autobiographic elaboration of characteristics and features of IS [19]. The cognitive processing of IS characteristics is especially important for the final product evaluation and choice as well as the justifiability of product choices. For example, recent studies have shown that the processing of pragmatic and hedonic product characteristics results in a cognitive bias. Although users appreciate hedonic product characteristics in terms of positive experiential outcomes, these characteristics are not valued in choice situations because pragmatic choices are easier to justify than hedonic choices [18]. This bias of justifiability is closely related to the construct of cognitive dissonance. For example, cognitive dissonance arises when the context of use rewards external instrumental outcomes, whereas the actual use is motivated intrinsically or results in experiential outcomes [21]. A cognitive strategy to reduce cognitive dissonance is to overlook the pleasurable outcomes and attribute instrumental outcomes to the IS usage. This rational process can be described with the following cognition: "1 am voluntarily spending a lot of time on this and enjoying it, therefore, it must be useful." [21: p. 676].

In general, the cognitive basis of IS affordances highlights that detecting and using IS features creates mental workload. For example, mental workload is created by the comparison of actual IS characteristics with the users' expectations and mental anchors for these characteristics (i.e., Design-Expectation Fit, Anchoring Theory, Expectation-Confirmation Model) [e.g., 35]. Also closely connected to mental workload is the construct of cognitive absorption. Cognitive absorption characterizes a state of total attention, in which lots of cognitive resources are allocated to using a specific IS [e.g., 21]. For example, Lowry et al. [34] integrated the single second-order constructs of the first-order construct cognitive absorption, namely control, curiosity, heightened enjoyment, immersion and temporal dissociation as intrinsic motivators into their HMSAM. This integration helped to further enhance the predictive validity and conceptual understanding of intrinsically motivated IS use.

3.3 Affective Perspective on IS Affordances

In total eleven papers focused on an affective perspective on IS affordances. Affective theories are receiving greater attention since recent studies on IS adoption have shown that emotions are a considerable result of a users' interaction with IS [e.g., 5, 25]. One example of affective reactions to IS usage is provided by Codish and Ravid [25]. The authors implemented cognitive and gamified design principles in the educational context and demonstrated the effect of playfulness as a positive affective response to IS usage.

Another example is provided by Wang and Scheepers [5] in the Model for Hedonic Information System (HIS) Acceptance. Here, the authors identified three overlapping conceptual identities of users of hedonic IS. These identities are described as the computer user, the hedonic consumer and the player. The computer user is associated with the technology acceptance model [45] and the hedonic consumer is associated with the Hedonic Theory [46] from consumer behavior research. The player role is associated with two affective theories, namely the Pleasure-Arousal-Dominance (PAD) Theory [47] and Flow Theory [48]. The PAD Theory is also known as the Three-Factor-Theory of Emotion and states that affective reactions can be described by three main emotions, namely pleasure, arousal and dominance. The Flow Theory describes flow as a state of intense pleasure and involvement in a certain action. Similar to cognitive absorption, flow is associated with attention focus, perceived control, curiosity, and intrinsic interest. Based on their results, the authors even argue that the intrinsic motivators "emotional responses, imaginal responses, and flow experience are three main predictors of HIS acceptance" [5: p. 255].

3.4 Social Perspective on IS Affordances

Only seven of our identified papers considered a social perspective on IS affordances. Social affordances of IS mainly rely on three assumptions, namely (1) that users can personally relate to IS, (2) that users tend to interact with IS in a similar manner as in human-to-human relationships, and (3) that IS can also include the user in collective actions. Here, the IS can serve as a mediator between different users or the system and the user can even work together on a particular task [see 38].

Two examples of applying social affordances to IS are provided by Gnewuch et al. [20] and Oinas-Kukkonen [49]. For example, Oinas-Kukkonen [49] highlighted the importance of considering socio-psychological design principles. In the context of behavioral change support systems, the authors suggested that peoples' behavior can be influenced by persuasive IS through integrated social influence (i.e., social comparison, normative influence, and social learning). Here, for example, health and healthy lifestyles are promising fields of application of behavioral change support systems. The second example was provided by Gnewuch et al. [20] and concentrated on conversational agents for customer service. In this study, the authors turned the cooperative principle of conversation and the central assumptions of the social response theory into design principles.

3.5 Framework of Theories and IS Use Contexts

Figure 2 aggregates the findings described above into one framework that structures the selection of theories according to the IS use context (i.e., utilitarian, dual-purposed, and hedonic). The IS use context can be seen as a continuum that ranges from utilitarian IS to hedonic IS with dual-purposed use a hybrid of these two poles [3]. As explained above, the affordances of utilitarian IS mainly rely on the use of extrinsic motivators and hygiene factors. These factors are for example covered by classical technology acceptance models like the Technology Acceptance Model (i.e., TAM [45]) and the Unified Theory of Acceptance and Use of Technology (i.e., UTAUT [50]) [4]. In the context of hedonic IS, however, these models are no longer sufficient because these models lack the detailed integration of intrinsic motivators and predictors related to experiential and hedonic outcomes. A recent analysis [51] of the applications and extensions of UTAUT has shown that among the many extensions of UTAUT, only two extensions focused on hedonic performance expectancy [52] or hedonic motivation [53]. However, even these UTAUT extensions only regard hedonic components as side effects and do not set the focus on hedonic components [54]. This is why newer models that focus on triggering the user's intrinsic motivation like for example the Hedonic-Motivation System Adoption Model (HMSAM) [34] have to be taken into account in this context. The affordances of dual-purposed systems rely on a combination of the theoretical basis of utilitarian and hedonic IS. For example, the Multimotive Information Systems Continuance Model (MISC) [35] is based on the distinction and combination of intrinsic and extrinsic motivators [40].

4 Discussion, Future Research and Conclusion

This literature review contributes to the understanding of the affordances of utilitarian, hedonic, and dual-purposed IS by providing an overview of theoretical perspectives that can be used for the creation, application, and evaluation of theory-based design principles. Our review highlights that a variety of scientific disciplines including IS,

Technology Type	Theoretical Perspective			
	Motivational	Social	Cognitive	Affective
Utilitarian	27b, 29, 30b, 33	3	13, 14, 17, 24, 25, 35, 37, 38, 40a, 43, 44, 45	13, 14, 17, 24, 25
Dual-Purposed	2, 16, 27, 28, 30, 31, 32	1, 2, 3, 4, 5, 6, 7, 8	5, 10, 12, 13, 14, 17, 21, 24, 25, 34, 35, 36, 37, 38, 39, 40b, 41, 42, 43, 44, 45, 46	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25
Hedonic	2, 16, 26, 27a, 28, 30a	1, 2, 4, 8	12, 17, 19, 20, 21, 35, 36, 38, 39, 41, 42, 45, 46	9, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23

Figure 2. Detailed Framework: Classification of theoretical basis according to the context of IS use (*/# = theories that were grouped into more than one theoretical perspective/more than one technology type; numbers refer to the numbers given to the theories in Figure 1)

behavioral economics, human-computer interaction and psychology contain basic research, theories and models that can be used to derive IS affordances. Our findings show that previous research on IS affordances can be grouped according to the four main theoretical perspectives: (1) motivational, (2) cognitive, (3) affective, and (4) social. Among these perspectives, motivational affordances can be seen as higher order affordances that can be translated into IS characteristics and features through cognitive, social and affective affordances. For example, intrinsic motivators are often conceptualized as emotions like fun, enjoyment, playfulness, pleasure, arousal or dominance [5] and motivational needs are often conceptualized as psychological (i.e., autonomy, competence, and relatedness) or social needs (i.e., achievement, affiliation and intimacy, and leadership and followership) [38]. However, the development and application of these theories in the context of IS design reveal some shortcomings which should be addressed in future research (see Figure 3). Our analysis illustrates that we need to learn more about the correct application and modification of existing theories from IS, behavioral economics, human-computer interaction and psychology in the context of motivational, cognitive, affective, and social affordances of future IS. The artefact of this literature review provides a basis to use existing interdisciplinary theories and models systematically to create, apply, and evaluate IS affordances and their impact on users. Before inventing new grounded theory for the affordances of utilitarian, hedonic, and dual-purposed technologies, we need to reinvent existing theories, i.e., extending them among motivational and hedonic components.

The first research gap is that motivational affordances are simply underutilized [38, 57]. Except of some positive examples mentioned above [e.g., 41, 42], there is still a need for more applications and evaluations of motivational affordances. Precisely, "it would be useful as a next research step to prototype and isolate design features that are intended to fit certain task motivations and expectations" [35: p. 539] and thereby isolate single effects and deepen our understanding of the effects of applied motivational affordances. Here, it would be interesting to compare the effects and predictive power of certain extrinsic and intrinsic motivators in distinct usage contexts (i.e., utilitarian vs. hedonic vs. dual-purposed). Furthermore, if applied, most design principles for motivational affordances are very high level, not context-sensitive and not on feature level [e.g., 57]. User experience is, however, very sensitive to the context in which a technology is used [58–60]. Therefore, there is a need for more context-specific evaluation of lower level applications of design principles of motivational affordances.

The second recommendation for future research considers the context of applying and evaluating motivational affordances. Existing applications and evaluations of design principles of motivational affordances can mainly be found in the context of gamification or gamified systems [e.g., 5, 33, 38]. There is, however, a need to study the application of motivational affordances in the context of less hedonic dual-purposed user assistance systems. Dual-purposed user assistance systems are mainly used voluntarily and during leisure time. Their purpose is, however, not only to enhance the users' enjoyment but also enhance their individual instrumental outcomes. Here, it would be interesting to compare the effects of certain motivational affordances in form of intrinsic motivators in this two usage contexts, namely (1) motivational affordances in gaming and gamified systems and (2) motivational affordances in non-gamified systems or rather less hedonic dual-purposed systems. This research agenda would also contribute to the present debate about the effect of gamification in non-gaming applications [61]. For example, in the context of cognitive and behavioral decision theories, it would be interesting, to investigate how the presentation of information (i.e., designed according to motivational affordances vs. purely pragmatic design) influence cognitive processes like decision-making or elaboration.

The third issue this review has identified, concerns the methods used and outcome variables measured to evaluate implementations of motivational affordances. For example, Wu and Lu [3] found that the relevance of different intrinsic and extrinsic motivators varied depending on the considered dependent variable in the research model. Consequently, future studies should always consider multiple outcome variables and pay attention to the possible differences in terms of predictive validity of the considered antecedents. Besides integrating multiple outcome variables, there is a need to combine multiple measures in order to prevent common-method bias [62]. Nearly all identified studies used self-report measures. Focusing on explicit measures exclusively might, however, result in an incomplete picture of the outcomes of motivational affordances. Therefore, implicit or rather unconscious antecedents should be studied. Here, integrating neurophysiological measures (e.g., electroencephalography) is a promising research field [26].

Finally, the fourth research gap concerns the personality of the user. Recent research has demonstrated that applied motivational affordances are perceived differently depending on the personality traits of the user [33, 38]. Different types of users prefer to use different motivating IS. Hence, “[a]pplications designed to accommodate multiple experience tracks for different personality traits could contribute to the sustained use of the application and enable users to better meet their personal goals” [33: p. 82]. A recent example of considering personality in the acceptance of dual-purposed IS was presented by Oettl, Berger, Böhm, Wiesche and Krcmar [63]. The authors classified six archetypes of users of enterprise social networks based on the two dimensions individual openness and perceived task-fit. In the context of motivational affordances in the consumer context, we need similar archetypes based on a combination of personality and motivational affordances.

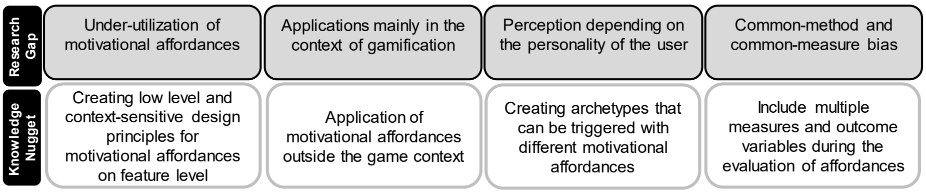


Figure 3. Summary of Identified Research Gaps and Potential Knowledge Nuggets

In sum, although motivational affordances should be a key requirement for IS, many ISs are not based on grounded theories and empirical insights on human motivation, cognition, affective reactions, and social interactions. Our review highlights that there is no need to invent new grounded theory on IS affordances. Instead, we need to rethink existing theories. Therefore, future research should apply and modify the identified theories and models from IS, behavioral economics, human-computer interaction and psychology in order to derive theory-based affordances for IS design. Since past research has mainly focused on the gaming context and gamified elements are more and more used in non-gaming applications, our research agenda focuses on motivational affordances in the context of non-gamified and dual-purposed systems. Here, low level and context-sensitive design principles for motivational affordances on feature level are needed. This is especially important for dual-purposed IS which should combine intrinsic and extrinsic motivators. For the evaluation of these design principles it is important to include multiple measures and outcome variables in order to avoid common-method bias and biases related to a specific outcome variable. Moreover, the interplay of personality traits and motivational affordances should be further studied in order to create archetypes that can be triggered with different motivational affordances. In sum, “taking into account a user’s motivational needs is one of the most crucial (but often neglected) design aspects for IS” [38: p. 271].

References

1. Delone, W., Mclean, E., Petter, S., Mclean, E.R.: The Past, Present, and Future of IS Success. J. Assoc. Inf. Syst. 13, 341–362 (2012)
2. Liu, D., Santhanam, R., Webster, J.: Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. MIS Q. 41, 1011-1034 (2017)

3. Wu, J., Lu, X.: Effects of Extrinsic and Intrinsic Motivators on Using Utilitarian, Hedonic, and Dual-Purposed Information Systems: A Meta-Analysis. J. Assoc. Inf. Syst. 14, 153–191 (2013)
4. van der Heijden, H.: User Acceptance of Hedonic Information Systems. MIS Q. 28, 695–704 (2004)
5. Wang, Z., Scheepers, H.: Understanding the Intrinsic Motivations of User Acceptance of Hedonic Information Systems: Towards a Unified Research Model. Commun. Assoc. Inf. Syst. 30, 255–274 (2012)
6. Gartner, https://www.gartner.com/doc/1528016/gamification-primer-life-game (Accessed: 28.06.2018)
7. Forbes, https://www.forbes.com/sites/michellegreenwald/2014/09/15/gamification-in-everything-the-range-and-when-and-why-its-so-effective/#787ff5936d32 (Accessed: 28.06.2018)
8. Arstechnica, https://arstechnica.com/features/2012/02/gamification-green-tech-makes-energy-use-a-gameand-we-all-win/ (Accessed: 28.06.2018)
9. Thong, J.Y.L., Hong, S.-J., Tam, K.Y.: The Effects of Post-adoption Beliefs on the Expectation-confirmation Model for Information Technology Continuance. Int. J. Hum. Comput. Stud. 64, 799–810 (2006)
10. Kakar, A.K.: Why do users speak more positively about Mac OS X but are more loyal to Windows 7? Comput. Human Behav. 44, 166–173 (2015)
11. Brauer, B., Ebermann, C., Kolbe, L.M.: An Acceptance Model for User-Centric Persuasive Environmental Sustainable IS. In: Proceedings of the Thirty Seventh International Conference on Information Systems, pp. 1–21. AIS, Dublin, Ireland (2016)
12. Hamari, J., Koivisto, J., Sarsa, H.: Does gamification work? - A literature review of empirical studies on gamification. In: Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 3025–3034. IEEE Computer Society Press, Waikoloa, HI, USA (2014)
13. CMO by Adobe, http://www.cmo.com/features/articles/ 2012/10/24/game-over-for-gamification.html#gs.m=glVps (Accessed: 28.06.2018)
14. Enterprise Gamification Consultancy, http://enterprise-gamification.com/mediawiki/index.php?title=Disneyland_Leaderboard_for_House_Keeping_Staff (Accessed: 28.06.2018)
15. Gartner, https://www.gartner.com/newsroom/id/2251015 (Accessed: 04.09.2018)
16. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, 13–23 (2002)
17. Senkbeil, M., Ihme, J.M.: Motivational Factors Predicting ICT Literacy: First Evidence on the Structure of an ICT Motivation Inventory. Comput. Educ. 108, 145–158 (2017)
18. Diefenbach, S., Hassenzahl, M.: The Dilemma of the Hedonic – Appreciated, but Hard to Justify. Interact. Comput. 23, 461–472 (2011)
19. Bhandari, U., Neben, T., Chang, K., Chua, W.Y.: Effects of Interface Design Factors on Affective Responses and Quality Evaluations in Mobile Applications. Comput. Human Behav. 72, 525–534 (2017)
20. Gnewuch, U., Morana, S., Maedche, A.: Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: Proceedings of the International Conference on Information Systems (ICIS), pp. 1–13. AIS, Dublin, Irland (2017)
21. Brade, J., Lorenz, M., Busch, M., Hammer, N., Tscheligi, M., Klimant, P.: Being there Again – Presence in Real and Virtual Environments and its Relation to Usability and User Experience Using a Mobile Navigation Task. Int. J. Hum. Comput. Stud. 101, 76–87 (2017)

22. Choudrie, J., Dwivedi, Y.K.: Investigating Broadband Diffusion in the Household: Towards Content Validity and Pre-Test of the Survey Instrument. In: Proceedings of the Thirteenth European Conference on Information Systems, pp. 38-57. AIS, Regensburg (2005)
23. Ebermann, C., Piccinini, E., Busse, S., Leonhardt, D., Kolbe, L.: What Determines the Adoption of Digital Innovations by Digital Natives? – The Role of Motivational Affordances. In: Proceedings of the Thirty Seventh International Conference on Information Systems, pp. 1–21. AIS, Dublin (2016)
24. Hassenzahl, M., Ullrich, D.: To Do or Not to Do: Differences in User Experience and Retrospective Judgments Depending on the Presence or Absence of Instrumental Goals. Interact. Comput. 19, 429–437 (2007)
25. Hassenzahl, M., Diefenbach, S., Göritz, A.: Needs, Affect, and Interactive Products – Facets of User Experience. Interact. Comput. 22, 353–362 (2010)
26. Hassenzahl, M., Schöbel, M., Trautmann, T.: How Motivational Orientation Influences the Evaluation and Choice of Hedonic and Pragmatic Interactive Products: The Role of Regulatory Focus. Interact. Comput. 20, 473–479 (2008)
27. Helfenstein, S.: Increasingly Emotional Design for Growingly Pragmatic Users? A Report from Finland. Behav. Inf. Technol. 31, 185–204 (2012)
28. Jia, J.S., Jia, J., Hsee, C.K., Shiv, B.: The Role of Hedonic Behavior in Reducing Perceived Risk: Evidence From Postearthquake Mobile-App Data. Psychol. Sci. 28, 23–35 (2017)
29. Magni, M., Taylor, M.S., Venkatesh, V.: 'To Play or not to Play': A Cross-temporal Investigation Using Hedonic and Instrumental Perspectives to Explain User Intentions to Explore a Technology. Int. J. Hum. Comput. Stud. 68, 572–588 (2010)
30. Malhotra, A., Gosain, S., Hars, A.: Evolution of a Virtual Community: Understanding Design Issues Through a Longitudinal Study. In: Proceedings of the Eighteenth International Conference on Information Systems, pp. 59–74. AIS, Atlanta (1997)
31. Weiser, P., Bucher, D., Cellina, F., De Luca, V.: A Taxonomy of Motivational Affordances for Meaningful Gamified and Persuasive Technologies. In: Proceedings of EnviroInfo and ICT for Sustainability 2015, pp. 271–280. Atlantis Press, Copenhagen (2015)
32. Herzberg, F.: Work and the Nature of Man. World Publishing, New York, USA (1966)
33. Deci, E.L.: Intrinsic Motivation. Plenum Press, New York, USA (1975)
34. Lowry, P.B., Gaskin, J.E., Twyman, N.W., Hammer, B., Roberts, T.L.: Taking Fun and Games Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM). J. Assoc. Inf. Syst. 14, 617–671 (2013)
35. Lowry, P.B., Gaskin, J.E., Moody, G.D.: Proposing the Multimotive Information Systems Continuance Model (MISC) to Better Explain End- User System Evaluations and Continuance Intentions. J. Assoc. Inf. Syst. 16, 515–579 (2015)
36. Resatsch, F.: Developing and Evaluating Ubiquitous Computing Applications Using the Example of Near Field Communication in German. Technische Universität München, München (2009)
37. Füller, K.J.: Customer Integration 4.0: Leveraging Customer Knowledge in Innovation Processes by Using Digital Technologies. Technische Universität München, München (2017)
38. Ajzen, I., Fishbein, M.: Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ. (1980)
39. Ajzen, I.: From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl, J. and Beckman, J. (eds.) Action-control: From Cognition to Behavior, pp. 11–39. SSSP Springer Series in Social Psychology, Heidelberg (1985)
40. Agarwal, R., Karahanna, E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. MIS Q. 24, 665–694 (2000)

41. Chai, J.C.Y., Malhotra, N.K., Alpert, F.: A Two-dimensional Model of Trust–Value–Loyalty in Service Relationships. J. Retail. Consum. Serv. 26, 23–31 (2015)
42. Codish, D., Ravid, G.: Personality Based Gamification: How Different Personalities Perceive Gamification. In: Proceedings of the European Conference on Information Systems, AIS, pp.1-11. Tel Aviv (2014)
43. Davis, F.D.: A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Massachusetts Institute of Technology, Cambridge (1986)
44. Hirschman, E.C., Holbrook, M.B.: Hedonic Consumption: Emerging Concepts, Methods and Propositions. J. Mark. 46, 92–101 (1982)
45. Russell, J.A., Mehrabian, A.: Evidence for a Three-Factor Theory of Emotions. J. Res. Pers. 11, 273–294 (1977)
46. Csikszentmihalyi, M.: Beyond Boredom and Anxiety. Jossey-Bass, San Francisco, USA (1975)
47. Oinas-Kukkonen, H.: Behavior Change Support Systems: A Research Model and Agenda. In International Conference on Persuasive Technology, pp. 4-14. Springer, Berlin Heidelberg (2010)
48. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D.: User Acceptance of Information Technology: Toward a Unified View. MIS Q. 27, 425–478 (2003)
49. Venkatesh, V., Thong, J.Y.L., Xu, X.: Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. J. Assoc. Inf. Syst. 17, 328–376 (2016)
50. Lallmahomed, M.Z.I., Ab.Rahim, N.Z., Ibrahim, R., Rahman, A.A.: Predicting Different Conceptualizations of System Use: Acceptance in Hedonic Volitional Context (Facebook). Comput. Human Behav. 29, 2776–2787 (2013)
51. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Q. 36, 157–178 (2012)
52. Rocznik, D., Goffart, K., Wiesche, M.: Designing Hedonic User Experiences: The Effect of Psychological Need Fulfilment on Hedonic Motivation. In: Proceedings of the Twenty-Fifth European Conference on Information Systems. pp. 3223–3233. AIS, Guimares (Portugal) (2017)
53. Zhang, P.: Motivational Affordances: Reasons for ICT Design and Use. ACM Commun. 51, 145–147 (2008)
54. Law, E.L.-C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., Kort, J.: Understanding, Scoping and Defining User Experience: A Survey Approach. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09, pp. 719–728. Association for Computing Machinery, Boston (2009)
55. Lin, C.-P., Bhattacherjee, A.: Extending Technology Usage Models to Interactive Hedonic Technologies: A Theoretical Model and Empirical Test. Inf. Syst. J. 20, 163–181 (2010)
56. Venkatesh, V., Thong, J.Y.L., Chan, F.K.Y., Hu, P.J.-H., Brown, S.A.: Extending the Two-stage Information Systems Continuance Model: Incorporating UTAUT Predictors and the Role of Context. Inf. Syst. J. 21, 527–555 (2011)
57. Karanam, Y., Alotaibi, H., Filko, L., Makhsoom, E., Kaser, L., Voida, S.: Motivational Affordances and Personality Types in Personal Informatics. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pp. 79–82. UbiComp '14 Adjunct, Seattle (2014)
58. Treiblmaier, H., Putz, L.-M., Lowry, P.B.: Setting a Definition, Context, and Theory-Based Research Agenda for the Gamification of Non-Gaming Applications. Assoc. Inf. Syst. Trans. Human-Computer Interact. 10, 129-163 (2018)

59. Podsakoff, P.M., Organ, D.W.: Self-Reports in Organizational Research: Problems and Prospects. J. Manage. 12, 531–544 (1986)
60. De Guinea, A.O., Titah, R., Léger, P.M.: Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. J. Manag. Inf. Syst. 30, 179–210 (2014)
61. Oettl, C., Berger, T., Böhm, M., Wiesche, M., Krcmar, H.: Archetypes of Enterprise Social Network Users. In: 51st Hawaii International Conference on System Sciences (HICSS), pp. 2036–2045. IEEE Computer Society Press, Manao, HI, USA (2018)

Eliciting Customer Preferences for Shopping Companion Apps: A Service Quality Approach

Tobias Wulfert[1], Jan H. Betzing[2], and Jörg Becker[2]

[1] University of Duisburg-Essen, Essen, Germany
tobias.wulfert@icb.uni-due.de
[2] University of Münster, ERCIS, Münster, Germany
{jan.betzing|joerg.becker}@ercis.de

Abstract. Shopping companion apps, which assist customers in product search and buying decisions, are an emerging phenomenon in the context of omnichannel retail. These retailer-provided apps link the digital with the physical servicescape of the store, allowing for new forms of online and at the same time physical service. So far, there is no dominant design for this type of information system. Both academia and practice lack empirical information about what customers expect from this kind of mobile app. Drawing from service quality literature as theoretical foundation, we conducted a qualitative content analysis of 1,448 customer reviews of three major shopping companion apps. The analysis yielded 23 aspects that customers expect from shopping companion apps, and that, in turn, can support establishing high mobile service quality. Our results contribute to the knowledge of m-service in retail and quality-driven app design.

Keywords: Mobile Service, Omnichannel, Service Quality, Customer Reviews.

1 Introduction

The mobile channel is an intimate and direct way for retailers to fulfill their customers' needs [1, 2] since customers nowadays use myriads of mobile devices in all situations of their daily lives including shopping [3, 4]. Customers search for products and services, and shop on the move without temporal or spatial constraints [5]. Customers are already using their smartphones in-store to get product information and compare prices [6]. Brick and Mortar (BaM) retailers respond to this changing customer behavior by introducing what we term *shopping companion apps* [1, 7].

A shopping companion app is conceptualized as retailer-provided software, executed on the customer's smartphone, which complements the personal and e-service of a BaM retailer by an additional mobile channel—manifesting as a digital support to the shopping process within and outside the store. Customers use these apps to access general shopping features such as product search, and immersive, location-based, and personalized functions such as in-store navigation and product recommendations [3, 8]. In contrast to third-party apps, shopping companion apps keep the customer in the retailers' self-contained environments [9] and allow them to create "seamless omni-

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

channel experience[s]" [10, p. 68] that match their overall strategies. Shopping companion apps offer several novel features of hybrid customer interaction [11] and value-added service [7], which are neither usefully realizable just in stationary retail nor in e-commerce [1, 12].

Being an emergent phenomenon, shopping companion apps have not yet received much academic investigation [1], and also instances in practice do not follow a dominant design [13] but rather vary in form and behavior. Developing and designing shopping companion apps that meet or even exceed the customers' expectations is a challenging endeavor [14]. To establish a high-quality shopping companion app as a mobile interface for customer interaction with the retailer, designers must consider smartphone-specific constraints (e.g., display size, mobile Internet) and opportunities (e.g., sensor access, immersion) atop traditional user experience aspects [15, 16].

We draw from Service Quality (SQ) research to identify what makes up a *high-quality* shopping companion app. SQ is concerned with assessing the quality of interactions between a customer and a service provider [17, 18] and is defined as the degree of "discrepancy between customers' expectations and perceptions" [19, p. 111] towards a received or experienced service [20]. SQ assessments have a long and rich history, ranging from person-to-person service (SERVQUAL) [17, 18], over e-service accessed via personal computers (E-S-QUAL) [21], to m-service accessed via mobile devices (M-S-QUAL) [22, 23]. Recent conceptualizations of Mobile App Service Quality (MASQ) now consider the peculiarities of mobile apps such as immersive human-computer interaction, location independence, and potentially far-reaching access to personal information and sensor data through the service provider [24-27]. However, while the literature acknowledges the importance of high-quality in-store service through the mobile channel as a future competitive edge for retailers [28], there is a lack of empirical information about what customers expect, and also–to the best of our knowledge–there is no domain-specific research for shopping companion apps.

Against this background, our research goal is *a set of customer preferences for shopping companion apps* with special consideration on MASQ. To reach this goal, we apply a twofold approach. First, we turn towards SQ literature as the theoretical foundation to understand MASQ. Second, we conduct a Qualitative Content Analysis (QCA) [29] of online customer reviews of three major shopping companion apps to retrieve the individual app users' subjective perceptions of SQ. We follow suggestions from human-computer interaction research to analyze the interaction with mobile apps from a user's point-of-view [16]. From the review corpus, we inductively derive 23 aspects customer prefer for shopping companion apps, which retailers can consider for quality-driven (shopping companion) app design.

The remainder of this paper is as follows: Section 2 introduces MASQ. Section 3 sketches the research approach. Section 4 gives the customer preferences for shopping companion apps, which are discussed in Section 5. We conclude in Section 6.

2 Theoretical Background

2.1 Service Quality

SQ is known to be an important determinant for the success of a company, impacting "business performance, lower costs, customer satisfaction, customer loyalty and profitability" [30, p. 913]. SQ assessments identify the perceived SQ of an individual, which is "a global judgment or attitude" [31, p. 16] comparing the customers' expectations and actual perceptions of a service endeavour. Following the so-called disconfirmation paradigm, high SQ is achieved, when the difference between expectations and perceptions is marginal, or the perceptions exceed the expectations [17]. In contrast, customer satisfaction is "the result of specific service transactions" [32, p. 822]. Existing models of SQ distinguish different sub-dimensions, which further detail the overarching construct. Parasuraman et al. [31], for example, introduced reliability, responsiveness, assurances, and tangibles as dimensions for measuring the quality of interpersonal service. Various technological innovations have led academia to propose adjusted and extended SQ models for different types of information systems, domains, and kinds of service over time. In the context of mobile apps and m-service, adjusted models are subsumed under MASQ. Currently, research on MASQ is sparse [26, 27].

SQ research has created an own literature stream that spreads over the information systems, retail, e-commerce, human-computer-interaction, and marketing domains. To identify the dimensions that may explain high-quality shopping companion apps, we conducted a structured literature review [33] across outlets in these domains, without restricting the search basket to allow for an exhaustive coverage. The search took place on 2017-06-21 using SCOPUS, AISeL, Web of Science, and EBSCOHOST. Table 1 gives the generalized search query, which was adapted to the syntax of the respective search engine.

Table 1. General Literature Review Search Query

```
(("app" OR "mobile" OR "electronic" OR "m-" OR "e-") AND
("service quality") AND ("*commerce" OR "*shopping") AND
("criteri*" OR "dimension" OR "measure*"))
```

After performing a one-way forward- and backward search, a sample of 34 papers remained, which yields insights on the determinants and dimensions of Electronic Service Quality (ESQ), Mobile Service Quality (MSQ), and MASQ in the B2C area.

2.2 Research Model of Mobile App Service Quality

Figure 1 shows the multidimensional, hierarchical research model of MASQ, adapted from [22]. Synthesized from the identified literature sample, the model comprises the dimensions of SQ, which apply to the case of shopping companion apps.

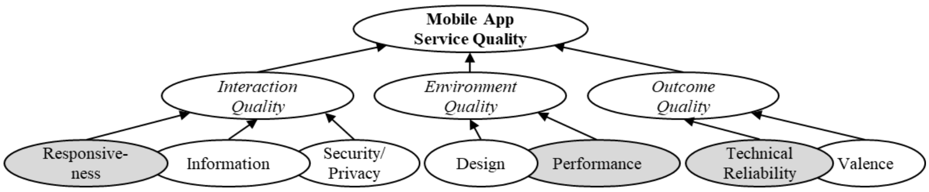


Figure 1. Multidimensional Hierarchical Research Model of MASQ, adapted from [22]

The shaded dimensions are either added to or reframed from the initial model by [22], based on the literature synthesis and initial results of the QCA. The example of the reliability dimension can illustrate this adaption. Initially, this dimension focuses on the reliability of delivery and fulfillment processes [34, 35], which are out-of-scope for shopping companion apps that are used primarily in-store. Reliability in the online customer reviews is often related to technical malfunctions and service dropouts. Hence, this dimension is subsequently denoted as technical reliability. In line with previous hierarchical models [36, 37], the MASQ is comprised of three secondary dimensions, which themselves are comprised by primary dimensions.

The research model in Figure 1 is used as the theoretical lens to analyze and structure the qualitative data in the subsequent QCA. Following the QCA terminology [29], the primary dimensions make up the *content categories*. To allow for a fine-grained analysis, these content categories are further detailed into 22 *characteristics*. Table 2 provides the content categories and characteristics related to MASQ, which have been derived from the literature review.

3 Research Approach

We analyzed customer reviews of three major, shopping companion apps by Walmart (US), Tesco (UK), and Marks & Spencer (M&S) (UK) to elicit customers' preferences for shopping companion apps. Although the selected retailers trade internationally, the apps under consideration are tailored to their respective home countries of operation. We sampled these apps because they are on the market for more than three years, they address a significant proportion of people in their countries of operation, and a vast number of online reviews make these three apps the subject of discussion. Further, these retailers have a long history in BaM operations. Additionally, we focused on grocery retail as industry because these retailers offer a broad spectrum of product categories, and thus, are potentially relevant for the majority of the population, compared to specialist shops that only address a certain focus group. Electronic customer reviews as a form of electronic word-of-mouth are a valuable source for indicating the quality of apps including the users' personal opinions, bug reports, and desired features [38]. We employ the QCA [29] as research design and make use of the methods presented in [39] and [40] to extract and prepare the review sample. The QCA approach is frequently applied to extract information from user-generated content such as reviews [41, 42].

Table 2. Content Categories and Characteristics related to Mobile App Service Quality

Dimension	Description	Support
Interaction quality	"Reflects all the quality characteristics of a customer's interaction with the [...] service provider." [22, p.942]	[22, 36, 37]
Responsiveness	The retailer's ability to promptly and politely solve a customer's issues related with the mobile app. **RES1:** Customer service availability **RES2:** Problem solving ability **RES3:** Politeness and kindness of personnel **RES4:** Guidance and instructions for app usage	[21, 35, 43–45]
Information	"The provision of accurate and precise information" [22, p.943] by the retailer. **INF1:** Information adequacy **INF2:** Information usefulness **INF3:** Information correctness	[22, 34, 35, 46–48]
Security and Privacy	"The protection of system and network resources from any external or internal attack and the protection of users' personal data." [22, p.943] **SEC1:** Information security **SEC2:** Data protection **SEC3:** Data collection	[21, 22, 35, 44, 45, 48]
Environment quality	Reflects "the context in which [mobile apps] are delivered, [and] quality characteristics of the equipment" [22, p.942] that affect the delivery of the mobile apps.	[22, 36, 37]
Design	The aesthetics, features, and layout of the user interface. **DES1:** Visual aesthetics and clarity of layout **DES2:** Quality of multimedia content **DES3:** Ease of use and ease of navigation **DES4:** Search function and filters	[21, 22, 34, 35, 45, 46, 48]
Performance	The performance of the mobile app and its resource requirements. **PERF1:** Processing speed **PERF2**: Device storage usage and mobile network usage **PERF3:** Network connection quality	[21–23, 34, 46, 49]
Outcome quality	Reflects the technical quality of and the customer's satisfaction with the service delivery.	[22, 36, 37]
Technical reliability	The accurate and consistent operation of the mobile app. **REL1:** Mobile app reliability **REL2:** Availability of provided services **REL3:** Continuous operation after updating	[22, 23, 45, 46, 48, 50]
Valence	The customer's ex-post impression of the service delivery. **VAL1:** Overall satisfaction with the provided service **VAL2:** Satisfaction with the scope of provided services	[22, 43, 46, 47, 51, 52]

Following Pagano and Maalej [39], we first extracted review data from Apple's App Store and Google's Play Store using a paid version of the online service *heedzy.com* on 2017-07-03. The raw data contains information on the app name, date, title, and content of the review, the nickname of the customer who created the review, the rating that is provided, and the app version. 10,099 reviews have been extracted in total. This number comprises 6,048 reviews for the Walmart app (1,084 iOS / 4,964 Android), 3,389 reviews for the M&S app (818 iOS / 2,571 Android), and 662 reviews for the Tesco app, where a technical restriction only allowed us to extract reviews from the Apple App Store. Since mobile apps are frequently updated, we only consider reviews written within the last six months, which leaves 8,237 reviews. The review sample was manually pruned by non-informational reviews such as "*Great app!*" and "*Useless!*".

Customer reviews in foreign languages and reviews concerning the retailer's general assortment, delivery quality, and price politics were excluded, which leaves a final sample of 1,448 reviews (795 Walmart, 433 M&S, 220 Tesco) for further investigation.

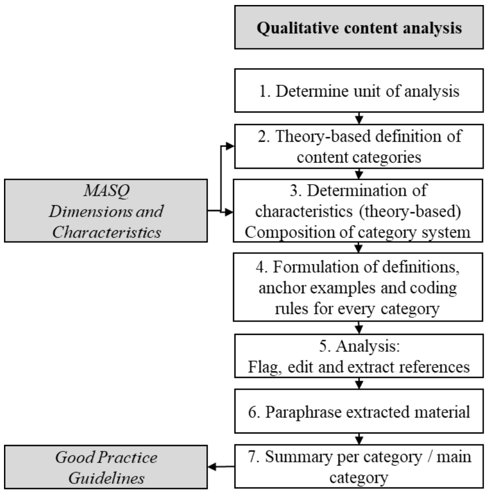


Figure 2. Research Approach, adapted from [29, 33]

Figure 2 shows the QCA process based on its content structuring approach, which supports the deductive assignment of reviews to content categories [29]. The review corpus containing the users' expectations and suggestions was grouped and analyzed for each content category. Anchor examples are used as references to illustrate the elicitation of customer preferences. Since online customer reviews usually contain more than one aspect [39], reviews were sub-classified regarding the characteristics, which results in 1,307 codings. We used *QCAmap.org* to aid the coding process [29].

4 Results

4.1 Relevance of MASQ Dimensions

We assigned the customers' statements to the content categories and characteristics, which allows weighting the relevance of the MASQ dimensions. Table 3 provides the distribution of customer review codings. The amount of codings per characteristic varies between five and 298. In effect, customers emphasize some of the dimensions and characteristics with higher importance than others. For example, *ease of use and ease of navigation* (DES3) as a part of the *design* dimension seems to be more important than attentive customer service (RES3) as a part of the *responsiveness* dimension. Regardless of the provided functionality, customers expect fast response times (PERF1) and reliable service (REL1) and use the review function of the app stores as an outlet to complain when issues arise.

Table 3. Frequency of Characteristics Mentioned in the Online Customer Reviews

Content Category	Σ	Mentions per Characteristic			
Responsiveness	93	RES1: 27	RES2: 52	RES3: 5	RES4: 9
Information	103	INF1: 29	INF2: 28	INF3: 46	
Security and Privacy	67	SEC1: 35	SEC2: 16	SEC3: 16	
Design	452	DES1: 38	DES2: 20	DES3: 298	DES4: 96
Performance	140	PERF1: 119	PERF2: 10	PERF3: 11	
Technical Reliability	394	REL1: 312	REL2: 19	REL3: 63	
Valence	59	VAL1: 45	VAL2: 13		

4.2 Customer Preferences Regarding Shopping Companion Apps

The three shopping companion apps under review provide a similar range of functions: Access to the particular online shop, click & collect, store finder, in-store inventory checking, aisle locator, promotions, loyalty programs, and product scanning. In the following, we focus on the non-functional aspects of app and service delivery. From the review corpus, we found rich hints what customers expect from shopping companion apps and identified 23 aspects that retailers could pick up to improve their m-service offerings. Table 4 lists the aspects, which we have structured by the primary dimensions of the MASQ model (Figure 1).

Customers voiced some preferences that are specific to the context of shopping companion apps. For example, because shopping companion apps enhance the physical servicescape of the store with digital service, customer service should be able to support both the retailer's m-service and issues that arise in-store. Nevertheless, trained customer service—of course—is also relevant to any other m-service. We highlighted the aspects that include such peculiarities regarding shopping companion apps in boldface. However, to provide a complete picture of the customers' preferences, we also include more general aspects in Table 4, which may fit other types m-service too.

Table 4. Customer Preferences for Shopping Companion Apps

#	Aspect	Description
Responsiveness		
01	**Trained customer service**	Customer service should be able to assist the customers knowledgeably and politely with any inquiries related to the retailer's mobile services, app functionality, and in-store issues.
02	**In-app guidance**	The app should include an onboarding process to introduce the retailer's range of provided m-service offerings and should provide a help section with usage instructions.
03	**Omnichannel customer service**	Customer service should be available through all channels the retailer offers and provide personalized service to customers independently of the selected channel.
04	In-app customer service	The app should provide direct access to customer service (e.g., by text, voice or video chat) and display easy to find contact information.
05	Responsive customer service	Customer service should respond timely to customer requests, even during times of high request volumes.
Information		
06	**Real-time information**	Any information shown in the app (e.g., prices, stock information) should be up-to-date, correct, complete, and consistent to information provided by the retailer through other channels.
07	Adequate and clear information	Any information shown in the app should be provided to the customer in a relevant, clear, and intelligible manner.
08	Update descriptions	Customers should receive detailed update and release notes.
Security and Privacy		
09	Request permissions	The customer's personal and payment data should be collected, stored, and processed only after permission for the particular purpose is granted.
10	Restrict permissions	Permissions should only be requested when they are required and appropriate for the app's provided set of functions.
Design		
11	**Limit advertisements**	Although customers expect to receive offers through shopping companion apps, advertisements should be used with moderation and not distract users from fulfilling their current objective.
12	**Accurate product search**	The app should include a product search engine that returns adequate results, which may also take the customer's current in-store context into account.
13	**Convenient product filters**	The app should provide filters to ease digital and physical information and product search processes.
14	High-quality multimedia content	Multimedia content such as product images and videos should be in high-quality and fit to the screen resolution.
15	Clear design and intuitive layout	The app should have a clean and simple design that fosters its Intuitive use.

#	Aspect	Description
16	Short navigation paths	All functions of the app should be easily accessible and not deeply nested.
Performance		
17	Technical responsiveness	The app should start up quickly and react fast to the customer's interactions.
18	Reasonable resource utilization	The app should have a small footprint regarding app and update sizes, local storage occupancy, and (mobile) network traffic.
19	Reduced background activity	The app should keep background activities to a minimum to limit battery drain.
Technical Reliability		
20	Reliable operations	The app should provide its service reliably without crashes or service outages.
21	Sustainable updates	Updates should be non-breaking and sustain existing functions.
Valence		
22	Scope of features and services	A single app should incorporate all m-service offerings that the retailer provides.
23	Continuous improvement	The retailer should use a continuous improvement process to react to ever-changing customer expectations.

Customer reviews frequently contain experiences on service encounters with the service provider. Customer service representatives constitute a direct personal contact within the otherwise human-to-machine context of shopping companion apps. Customers expect service personnel that will respond in a timely fashion when they require assistance, is reachable through the channel of their choice, and can quickly, professional, and politely resolve their issues. We found that customers, among other things, complain about a lack of contact channels and unavailable representatives ("*No one to call no one to email*"), and the quality of employee training ("*The employees are clueless on how to even help you*"). However, they also mention positive service experiences ("*Had problem just called and they had it fixed in 2 mins thanks*").

Especially for shopping companion apps whose information is changing at a fast pace (offers, stock information), users expect correct, current, and complete data that is relevant to their situation ("*Used every day to keep updated with offers and developments*"). Incorrect or useless information discourages users ("*If someone is using the app and looking for the store nearest to them, why would you have a distribution center come up as the closest store and then direct people there?*") and can result in lost sales.

Regarding the app itself, customers expect guidance on the use of the app ("*Can't find instructions on how to use, and there are features that are not all that intuitive*") and want to be informed of changes introduced by app updates ("*When YOU update your app, you need to specify what changed - features, bug fixes, etc. BS like WE MADE IT BETTER is not an update description*").

Similarly, the retailer has to be transparent on the collection and processing of personal data so that customers gain trust and grant the requested permissions ("*Major* *privacy concern I just used the app to make a purchase. Without my knowledge or permission, it stored my credit card information.*"). In effect, permissions and data should only be requested when they are required for the proper provision of the app's service and features ("*Unwarranted permissions - why does it need access to WiFi info, media files and photos, contacts - no, really M&S, and no explanation*").

Design and usability are the most frequently reviewed aspects in our sample (see Table 3). While app designers receive numerous hints of varying relevance from the reviews, we suggest executing a structured usability assessment [53]. Users particularly name aspects such as a clean and simple design ("*Good size text and clear with suitably neutral colours which are pleasing to the eye*"), intuitive navigation ("*The app is far quicker & easier to use than the website*"), high-quality multimedia content ("*doesn't get 5 stars because of the ui and graphics. Low resolution or not optimized for retine screens*"), a reliable search engine, and filtering functions ("*Why doesn't this app allow me to sort my results? Like price high - low?*").

Customers expect well-performing apps ("*Opening app takes a long time*"), which make reasonable use of cellular data ("*Too data intensive. [...] it is taking FOREVER to load [...]*"), on-device storage ("*With storage space at a premium on my phone, I am seriously reconsidering the necessity of having this app*"), and background processing activities ("*It is active in the background to such a degree that it is a constant drain on my battery*"). Although the preferences regarding technical reliability seem obvious and should be taken for granted, customers reported a significant amount of technical issues for the apps under consideration. App developers need to make sure that the range of offered m-services and features is functioning correctly without bugs ("*Nullpointer exception when trying to add anything to the basket*") and continues to work after updating ("*Still can't use scanner...every since November 2016 update*").

The valence dimensions subsume the subjective feelings and perceptions customers have after using the retailer's m-service. We suggest to regularly assess the customers' feelings towards the app, e.g., by using an in-app survey mechanism to identify the individual pain points. Lastly, the apps under review apply two different strategies. Either, all customer-facing m-service offerings of the retailer are bundled in a single app, or there are multiple apps that all fulfill a single purpose. A frequently named ex post evaluation was not to spread features across multiple apps but to follow the first strategy ("*You have so many apps doing different things! [...] Link them for easy use*").

5 Discussion

Informed by extant SQ knowledge and real-world customer reviews of three major shopping companion apps, we provide a set of 23 aspects customer expect from shopping companion apps, structured by a multidimensional hierarchical model on MASQ. Nevertheless, fulfilling these aspects is not a *sufficient* condition for high MASQ *per se*, although they are based on the individual customers' written expectations and perceptions. As with most design decisions, there is no one-size-fits all approach since customers may have different or even contradicting expectations of the same service. Nevertheless, the derived preferences provide a good picture of what the majority of customers desire with regards to shopping companion apps. While these aspects can aid app developers and support quality-driven app design, their implementation might be difficult. Inherent with using customer reviews as data source, we draw from customers' wishful thinking that may conflict with the retailers' economic and business capabilities.

Since the earliest conceptions of SQ, most models are based on the disconfirmation paradigm [18], which states that quality is the result of the comparison between perceived and expected performance [17]. However, a small difference in performance as the indicator for high SQ is subject to critique by some scholars [e.g., 54]. Following the conceptualization, high SQ is achieved as long as the customer's expectations are met or exceeded, even though these expectations might be meager. In our understanding, the SQ is high when the expectations of a large majority of customers are exceeded. Suppose we have a large sample, outliers having particularly low respective particularly high expectations rule themselves out. Consequently, we used a large sample of 1,448 customer reviews to capture a representative set of customer expectations. On a further note, we confirm Knote et al's [40] observation that users' perceptions and evaluations strongly depend on their current situations and previous experiences with other apps, which are taken as a reference point for comparisons.

The elicited customer preferences constitute a snapshot in time that reflects current customers' perceptions and foci. Over time, these foci may change as new technologies and services evolve. For this reason and the subjective nature of SQ, we did not state explicit design guidelines that retailers have to implement to achieve high MASQ.

As with any research, our work comes with some limitations. First, our results may be prone to selection bias because the same set of researchers performed the derivation of content categories and characteristics related to MASQ as well as the subsequent coding of the online customer reviews.

Second, we did not take into account the specific Graphical User Interface (GUI) components and structures defined by the different mobile operating systems. These design conditions may influence both the customers' service expectation prior to the app usage, and their perception when interacting with the shopping companion apps on a smartphone. However, the scope of and interface to the provided functionality of the samples apps only slightly differed between Apple iOS and Google Android.

Third, the sampled apps all focus on grocery retail, which may constrain the generalizability of the derived customer preferences for shopping companion apps that focus on other retail industries such as apparel or sporting goods to some extent.

Fourth, there are inherent limitations when dealing with online customer reviews in general and reviews of mobile apps in particular. As customers can post online reviews anonymously [39], no customer information is available on the analyzed set of reviews. Consequently, we cannot make a detailed statement whether our sample is representative. Further, within the corpus of seemingly authentic online customer reviews published by real customers, there can be spam and misleading reviews [55]. Due to the anonymity of reviewers, we cannot rule out *manipulation* by developers and app providers to praise their product. We have no means to identify fake reviews. However, these reviews usually tend to praise or condemn the app under review without going into detail. Since we are only interested in information on the service delivery, we can quickly discard non-informational reviews such as "Useless rubbish!"—being a fake review or not.

Moreover, the majority of customers only posts a single review, but a small number of customer rates a mobile app several times (up to nine times in our sample). By using the app over a longer period, users become familiar and might identify further aspects worth reviewing. Therefore, we kept those reviews in the corpus. Admittedly, the number of reviews per customers has to be treated carefully because a single customer also can write reviews using multiple nicknames, which biases results in small samples.

Customers tend to report remarkably satisfying or dissatisfying aspects involving exaggerations and generalizations when reviewing products and services [55]. Although we use single anchor examples for clarification, our preferences abstract from single reviews and always summarize the judgments of many customers, in line with [29], so that single exaggerations do not bias our analysis.

We collected the most current 10,099 online customer reviews of the three shopping companion apps across a time span of six months. During this period, the apps have received multiple updates. For simplification, we did not include version numbers in the QCA. Thus, comparisons across app versions are not possible, and changes in the perceived MASQ cannot be traced back to new app releases.

Lastly, the manual coding process follows a strict procedure [29]. However, it may be biased by personal opinions and subjective evaluations. A fully automated process for the QCA, ranging from the extraction of online customer reviews from the app stores to summarizing the results per category using advanced text mining tools or cluster analyses could help to alleviate this issue. Nevertheless, a manual process seems appropriate in the context of subjective online customer reviews that may contain ironic and ambiguous statements, which are hard to identify for automated tools [39, 55].

6 Conclusion and Outlook

M-commerce is a significant growth area for retailers, and mobile shopping companion apps are an emerging phenomenon that has not found much attention in academia so far. To remedy this situation and set the field, this paper first introduced a working definition for shopping companion apps. We built upon kernel theory from SQ and introduced a multidimensional, hierarchical research model of MASQ that comprises the content categories and characteristics relevant to service delivered through mobile apps. From analyzing close to 1,500 real-world customer reviews of three major shopping companion apps, we identified 23 aspects customer expect from this kind of app. Our results can aid service and app designers towards providing mobile service and shopping companion apps that exhibit high MASQ.

We investigated shopping companion apps from a customer's point-of-view. In future work, we will assess the overall service system that spans between retailers and customers and between groups of customers. Special consideration will be put on the interactions between the involved parties and on the co-creation of shopping experiences that are facilitated by m-services and apps. Finally, we follow the dual mission of design science to advance theory while developing and evaluating innovative IT artifacts for practice. The customer preferences elicited in this work will inform our overarching design science research project on m-service in BaM retail.

Acknowledgements

This paper was developed in the research project *smartmarket*[2], which is funded by the German Federal Ministry of Education and Research (BMBF), promotion sign 02K15A074. The authors thank the Project Management Agency Karlsruhe (PTKA).

References

1. Rudolph, T., Nagengast, L., Melanie, B., Bouteiller, D.: Die Nutzung mobiler Shopping Apps im Kaufprozess. Marketing Review St. Gallen 3(32), 42–49 (2015)
2. Hagberg, J., Sundstrom, M., Egels-Zanden, N.: The Digitalization of Retailing: an Exploratory Framework. International Journal of Retail & Distribution Management. 44, 336–368 (2016)
3. Inman, J.J., Nikolova, H.: Shopper-Facing Retail Technology: A Retailer Adoption Decision Framework Incorporating Shopper Attitudes and Privacy Concerns. Journal of Retailing 93(1), 7–28 (2017)
4. Pantano, E., Viassone, M.: Engaging consumers on new integrated multichannel retail settings: Challenges for retailers. Journal of Retailing and Consumer Services 25, 106–114 (2015)
5. Faulds, D.J., Mangold, W.G., Raju, P.S., et al.: The mobile shopping revolution: Redefining the consumer decision process. Business Horizons. 61, 323–338 (2018)
6. Gensler, S., Neslin, S.A., Verhoef, P.C.: The Showrooming Phenomenon: It's More Than Just about Price. Journal of Interactive Marketing 38, 29–43 (2017)
7. Groß, M.: Mobile shopping: A Classification Framework and Literature Review. International Journal of Retail & Distribution Management 43(3), 221–241 (2015)
8. Betzing, J.H., Hoang, A.-Q.M., Becker, J.: In-store Technologies in the Retail Servicescape. In: Multikonferenz Wirtschaftsinformatik. pp. 1671–1682. Lüneburg, Germany (2018) Taylor, D.G., Levin, M.: Predicting Mobile App Usage for Purchasing and Information Sharing. International Journal of Retail & Distribution Management 42(8), 759–774 (2014)
9. Rigby, D.: The Future of Shopping. Harvard Business Review 89(12), 65–76 (2011)
10. Nüesch, R., Alt, R., Puschmann, T.: Hybrid Customer Interaction. Business and Information Systems Engineering 57(1), 73–78 (2015)
11. McLean, G., Al-Nabhani, K., Wilson, A.: Developing a Mobile Applications Customer Experience Model (MACE)- Implications for Retailers. Journal of Business Research 85, 325–336 (2018)
12. Suárez, F.F.: Battles for technological dominance: an integrative framework. Research Policy 33(2), 271–286 (2004)
13. Khalid, H., Shihab, E., Nagappan, M., Hassan, A.: What do mobile app users complain about? a study on free ios apps. IEEE Software 32(3), 70–77 (2014)
14. Hassenzahl, M., Tractinsky, N.: User experience—a research agenda. Behaviour & Information Technology 25(2), 91–97 (2006)
15. Mendoza, A.: Mobile user experience: patterns to make sense of it all. Morgan Kaufmann, Waltham (2013)

16. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A Conceptual Model of Service Quality and Its Implications for Future Research. The Journal of Marketing 49(4), 41–50 (1985)
17. Grönroos, C.: A Service Quality Model and its Marketing Implications. European Journal of Marketing 18(4), 36–44 (1984)
18. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research. Journal of Marketing 58(1), 111–124 (1994)
19. Zeithaml, V.A., Parasuraman, A., Malhotra, A.: A Conceptual Framework for Understanding e-Service Quality: Implications for Future Research and Managerial Practice. Marketing Science Institute Report pp. 1–26 (2000)
20. Parasuraman, A., Zeithaml, V.A., Malhotra, A.: E-S-Qual: A Multiple-Item Scale for Assessing Electronic Service Quality. Journal of Service Research 7(3), 213–233 (2005)
21. Stiakakis, E., Petridis, K.: Developing and Validating a Multi-Criteria Model to Evaluate Mobile Service Quality. In: Martínez-López,F. (ed.) Handbook of Strategic e-Business Management, pp. 935–956. Springer-Verlag, Berlin Heidelberg (2014)
22. Huang, E.Y., Lin, S.W., Fan, Y.C.: M-S-QUAL: Mobile Service Quality Measurement. Electronic Commerce Research and Applications 14(2), 126–142 (2015)
23. Shankar, V., Kleijnen, M., Ramanathan, S., Rizley, R., Holland, S., Morrissey, S.: Mobile Shopper Marketing: Key Issues, Current Insights, and Future Research Avenues. Journal of Interactive Marketing 34, 37–48 (2016)
24. Lee, Y.E., Benbasat, I.: A Framework for the Study of Customer Interface Design for Mobile Commerce. International Journal of Electronic Commerce 8(3), 79–102 (2004)
25. Kumar, R.: A Proposed Scale of Assessing Mobile App Service. ELK Asia Pacific Journal of Marketing and Retail Management 8(1), 1–11 (2017)
26. Kuo,T., Tsai,G.Y., Lu,I.Y., Chang,J.S.: Relationships among Service Quality, Customer Satisfaction and Customer Loyalty: A Case Study on Mobile Shopping APPs. In: Proceedings of the 17th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS '16). pp. 1–8. Taipei, Taiwan (2016)
27. Grewal, D., Roggeveen, A.L., Nordfällt, J.: The Future of Retailing. Journal of Retailing 93(1), 1–6 (2017)
28. Mayring, P.: Qualitative Content Analysis. Theoretical Foundation, Basic Procedures and Software Solution. Beltz, Klagenfurt, Austria (2014)
29. Seth, N., Deshmukh, S., Vrat, P.: Service Quality Models: A Review. International Journal of Quality & Reliability Management 22(9), 913–949 (2005)
30. Parasuraman, A., Zeithaml, V.A., Berry, L.L.: A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality. Journal of Retailing 64(1), 12–37 (1988)
31. Jun, M., Yang, Z., Kim, D.: Customers' Perceptions of Online Retailing Service Quality and Their Satisfaction. International Journal of Quality & Reliability Management 21(8), 817–840 (2004)
32. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: on the Importance of Rigour in Documenting the Literature Search Process. In: 17th European Conference on Information Systems. pp. 2206–2217. ECIS '09, Verona, Italy (2009)
33. Collier, J.E., Bienstock, C.C.: Measuring Service Quality in E-Retailing. Journal of Service Research 8(3), 260–275 (2006)
34. Holloway, B.B., Beatty, S.E.: Satisfiers and Dissatisfiers in the Online Environment. Journal of Service Research 10(4), 347–364 (2008)
35. Lu, Y., Zhang, L., Wang, B.: A Multidimensional and Hierarchical Model of Mobile Service Quality. Electronic Commerce Research and Applications 8(5), 228–240 (2009)

36. Stiakakis, E., Georgiadis, C., Petridis, K.: Analyzing m-Service Quality Dimensions Using Multivariate Statistical Techniques. In: Proceedings of the 12th International Conference on Mobile Business (ICMB '13). Berlin, Germany (2013)
37. Genc-Nayebi, N., Abran, A.: A Systematic Literature Review: Opinion Mining Studies from Mobile App Store User Reviews. Journal of Systems and Software 125, 207–219 (2017)
38. Pagano, D., Maalej, W.: User Feedback in the AppStore: An Empirical Study. In: Proceedings of the 21st International Requirements Engineering Conference (RE '13). pp. 125–134. Rio de Janeiro, Brazil (2013)
39. Knote, R., Söllner, M.: Towards Design Excellence for Context-Aware Services: The Case of Mobile Navigation Apps. In: Proceedings of the 13th Internationale Tagung Wirtschaftsinformatik (WI'17). pp.121–135. St.Gallen, Switzerland (2017)
40. Giannakos, M.N.: Absolute Price as a Determinant of Perceived Service Quality in Hotels: A Qualitative Analysis of Online Customer Reviews. International Journal of Hospitality and Event Management 1(1), 62–80 (2014)
41. Spreer, P., Rauschnabel, P.A.: Selling with Technology: Understanding the Resistance to Mobile Sales Assistant Use in Retailing. Journal of Personal Selling & Sales Management 36(3), 240–263 (2016)
42. Bauer, H.H., Falk, T., Hammerschmidt, M.: eTransQual: A Transaction Process-based Approach for Capturing Service Quality in Online Shopping. Journal of Business Research 59(7), 866–875 (2006)
43. Blut, M., Chowdhry, N., Mittal, V., Brock, C.: E-Service Quality: A Meta-Analytic Review. Journal of Retailing 91(4), 679–700 (2015)
44. Wolfinbarger, M., Gilly, M.C.: eTailQ: Dimensionalizing, Measuring and Predicting Retail Quality. Journal of Retailing 79(3), 183–198 (2003)
45. Fassnacht,M., Koese,I.: Quality of Electronic Services. Journal of Service Research 9(1), 19–37 (2006)
46. Loiacono, E., Watson, R., Goodhue, D.: WebQual: An Instrument for Consumer Evaluation of Web Sites. Int. Journal of Electronic Commerce 11(3), 51–87 (2007)
47. Santos, J.: E-Service Quality: A Model of Virtual Service Quality Dimensions. Managing Service Quality: An International Journal 13(3), 233–246 (2003)
48. Madu, C.N., Madu, A.A.: Dimensions of e-Quality. International Journal of Quality & Reliability Management 19(3), 246–258 (2002)
49. Swaid, S.I., Wigand, R.T.: Measuring the Quality of e-Service: Scale Development and Initial Validation. Journal of Electronic Commerce Research 10(1), 13–28 (2009)
50. Tan,F.B., Chou,J.P.C.: The Relationship between Mobile Service Quality, Perceived Technology Compatibility, and Users' Perceived Playfulness in the Context of Mobile Information and Entertainment Services. International Journal of Human Computer Interaction 24(7), 649–671 (2008)
51. Vlachos,P.A., Giaglis,G., Lee,I., Vrechopoulos,A.P.: Perceived Electronic Service Quality: Some Preliminary Results From a Cross-National Study in Mobile Internet Services. International Journal of Human-Computer Interaction 27(3), 217–244 (2011)
52. Albert, W., Tullis, T.: Measuring the user experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann, 2 ed. (2013)
53. Buttle, F.: SERVQUAL: Review, Critique, Research agenda. European Journal of Marketing 30(1), 8–32 (1996)
54. von Hoffen, M., Hagge, M., Betzing, J.H., Chasin, F.: Leveraging social media to gain insights into service delivery: a study on Airbnb. Information Systems and e-Business Management 16(2), 247–269 (2018)

The Role of Early User Participation in Discovering Software – A Case Study from the Context of Smart Glasses

Benedikt Zobel[1], Karl Werder[2], Lisa Berkemeier[1], and Oliver Thomas[1]

[1] Osnabrück University, Information Management and Information Systems, Osnabrück, Germany
{benedikt.zobel,lisa.berkemeier,oliver.thomas}@uos.de
[2] University of Cologne, Cologne Institute for Information Systems, Cologne, Germany
werder@wiso.uni-koeln.de

Abstract. Smart glasses facilitate advanced user interaction and increase workplace efficiency through innovation. Yet, their capabilities rely on user-driven discovery of new software that harnesses its benefits. This study investigates user participation during the discovery of new software, leveraging this emergent technology. We investigate user participation during software product discovery, i.e. during early activities that precede classical development and design activities, through an in-depth longitudinal case study with two representative user organizations. The results suggest an evolutionary perspective toward the benefits of different types of user participation: 1) user as a source of information, 2) user as a co-creator, and 3) user as an innovator. Practitioners benefit from our lessons learned, validation and extension of software discovery toward the emergent technology, and recommendations to apply user-driven software discovery. We distill three lessons: evolving types of user participation, enhancing desirability through user participation, and carefully discovering software products for emergent technologies.

Keywords: Smart Glasses; Software Product; Software Discovery; Augmented Reality; User Participation

1 Introduction

When developing new software products, the literature presents different techniques, processes, and methods such as Scrum, XP or DSDM that help businesses to become more agile and increase development performance [1], [2]. While agile methods provide very little guidance related to early activities prior to programming [3], [4], scholars investigate the extension of these techniques with ideas from user participation practices such as user-centered design [5], [6]. Therefore, we need to better understand the role of user participation in these early activities, also called software product discovery [1]. Given the nature of innovation shifts from mechanical systems toward the software element of such products [7], the discovery of software products is an

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

emergent research field [5]. The associated term of (software) product discovery describes the early activities that precede classical development and design activities. The term originates from the pharmaceutical domain where it is often associated with the discovery of new drugs (e.g. [8]). In other domains, such as new product development, the product discovery phase is described as ideation stage [9]. In innovation management, the terms "fuzzy front-end" and "front-end innovation" have been introduced (cf. [10]). Irrespective of the domain, the objective is to reduce uncertainty at inception of a product development and design project (e.g., [10]). Here, the phase involves users early in order to identify high level needs of the user base and assure the desirability of the solution [11].

The literature presents us with different types of user participation. From a development perspective, we can distinguish three types of user participation [2], [12], [13]. First, and often associated with a traditional waterfall methodology, the user is a source of information [12]. Second, the user can be engaged and co-create the software product with the product team [14], [15]. Third, the user can become an innovating force behind new trends by utilizing toolkits [12]. While some organizations may not decide to involve users at all, other may include them in design reviews or include a user representative as part of the product team [16–18]. However, little is known about the interrelations of different types of user participation.

Smart glasses are an emergent and disruptive technology that can have sustainable impact on society and businesses. The lightweight devices (such as Google Glass or Vuzix M300) are characterized by their high mobility and mobile internet connection [19]. Contrary to other disruptive technologies, such as social media and mobile technology, smart glasses gain faster adoption by businesses than consumers [20]. Businesses see the potential of smart glasses in the way it changes the interaction between the computer and its users [21]. For example, smart glasses used in service and business processes assist employees in conducting tasks through the augmentation of their reality [22], i.e. enriching an individual's reality with further information through digital technology. Particular innovation potential lies with the software and services offered for it [7]. Due to limited experiences with these new devices and the corresponding software, we need to understand the technology's induced changes and influence on users during the discovery of new software products in order to mitigate risks of change requests or unaccepted software products.

This study investigates user participation during the software discovery phase in the context of an emergent technology. Our research objectives are: i) to investigate user participation during software discovery, ii) to understand the interrelations of different types of user participation, and iii) to identify lessons learned from an application of software discovery in a smart glasses project. Consequently, we formulate the following research question: *How can development teams facilitate user participation during the software product discovery for smart glasses?*

Practitioners benefit from our identified activities, lessons learned, and their applicability to software discovery. These can be used as guidelines to instantiate the process of discovering software products for smart glasses [4]. Our theoretical contribution focuses on the investigation of user participation during early activities within software development [5]. First, we adapt existing approaches to discovery by focusing on the participation of users when identifying software products for an emergent technology [1]. Second, we investigate how different forms of user participation benefit the desirability of a product vision during product discovery [23]. Third, we identify the evolutionary nature amongst three types of user participation and suggest a preliminary process model [24].

2 Theoretical Background

2.1 Software Product Discovery

The product discovery process describes the generation of ideas that precedes the development of application functions, leading to a product vision or product discovery plan [5], [1] (Figure 1). The objective of such a vision or plan is to assure the product's feasibility (e.g. [25]), viability (e.g. [9], [26]) and desirability (e.g. [14]). Hence, specific product, user, and team related concerns need to be addressed in five distinct phases. *Initialization* suggests the need for a new development project to assure clearly defined boundaries, the availability of key resources, and key stakeholder commitment [5]. *Product vision building* provides a common understanding of planned software products amongst the development teams [27], [28]. *User engagement* describes integrating the user as early as possible into the discovery process [28], [29]. The participation of users assures gathering actual user goals and user needs and therefore, leads to a desirable solution [1], [28]. *Requirements specification* identifies software needs from their initial idea, to later refined requirements through user feedback [27], [29]. A product backlog will be filled with elements from this requirements specification, the product vision building, and the user engagement [1]. Following, the phase *development and design* of drafts and low-fidelity prototypes simplifies the communication of the vision and helps to gain further insights from users and stakeholders through additional iterations (cf. [29]).

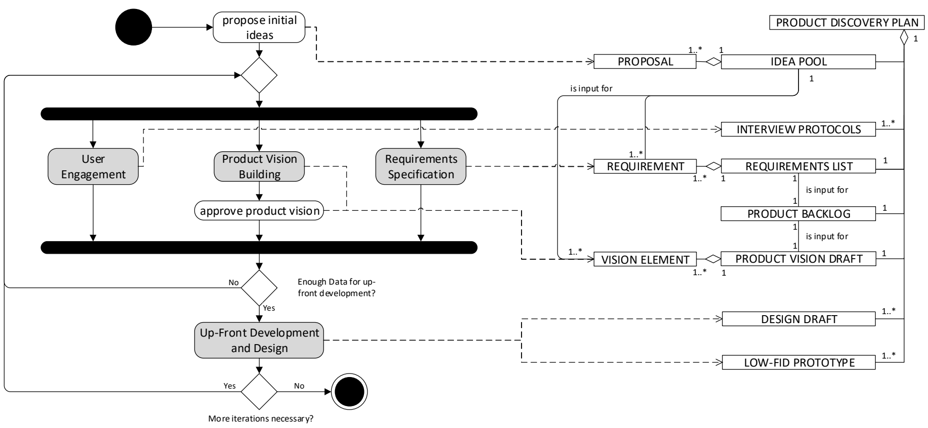


Figure 1. Software Product Discovery (simplified from Werder et al. [1])

2.2 User Participation

User participation has been an important factor in software development for decades [17]. Some of the benefits include quality enhancements, mitigation of unnecessary expenses, higher acceptance, increased use effectiveness, and enhanced development success (e.g. [2], [12]). While the literature presents differences between the concepts of user participation and a closely related concept of user involvement [30], recent studies suggest that scholars use these terms interchangeably most of the times [31], [32]. The positive effect between enabling participation of the user and system success has been empirically demonstrated [32–34]. Yet, system success has been measured in a variety of ways [31].

Prior studies suggest the multidimensionality of user participation [2], [15], [18]. One important dimension relates to the three different types of user participation [2], [12], [13]. First, the user can serve as a source of information. Designs are generated for the user. Second, the user can participate in the creation process as a co-creator. Hence, designs are generated with the user. Third, the user can be the innovating force behind new ideas, when designs are generated by the user. Despite the importance of user participation in software development, the role of user participation in early activities of the product development received little attention [5].

3 Research Method

Given the investigation of a contemporary phenomenon, the study adopts the case study research method [35]. The study presents a longitudinal single case with an embedded design in order to understand the changes in user participation during the discovery of software products for smart glasses. Given the quick adoption of smart glasses in businesses, the study focuses on the use of smart glasses to support employees in the logistics industry, e.g. through information provision or enabling user input during the execution of work-related tasks. The logistics industry is particularly interesting, as workers are confronted with many complex tasks established by manual labor, a field with high potential for technological support [36]. Following prior guidelines, the overall research design, the data collection and data analysis are presented in further depth [37]. Table 1 depicts the different activities and steps incorporating users in the instantiation and first iteration of the software discovery approach, including the duration in which the activities and phases took place. The activities are mapped to the main phases of product discovery (cf. [1], [10]). Since the early activities resulted in the design, development and evaluation of multiple prototypes, the varying timeframes of these iterations were documented in the table as range.

Table 1. Mapping of design technique to different discovery phases.

Discovery Phase / Design technique	*Duration (Months)*	*Expert Interviews*	*WH-Observations*	*Cost-Benefit-Discussion*	*Idea Consolidation*	*Product Vision Formulation*	*Benefit Evaluation*	*Requirements Formulation*	*Mockup Design & Prototyping (per iteration)*	*Prototype Evaluation (per iteration)*
Duration (Months)		1	3	1	2	1	1	1	1-3	1-2
Initialization	4	X	X							
User Engagement	11		X	X	X	X	X	X		X
Product Vision Building	4					X	X			X
Requirements Specific	4						X	X		X
Development & Design	4							X	X	

3.1 Research Design

The study follows a longitudinal single case design in order to investigate user participation during the software discovery process. The research context is the logistics industry, where new software for smart glasses is needed to support employees. The discovery phase guides the development of a product vision. Two prospective user organizations that are interested in the software products idea take part in this study. While company ALPHA, a large German logistics handler providing global services, is specialized on contract logistics with a staff of over 12,000 and more than 3,1 billion € annual turnover, company BETA is a German organization mainly offering fashion transport and services with 2,500 employees and more than 250 million € annual turnover. ALPHA's aim of the project is to relieve employees ergonomically and protect them from mistakes in work security. Additionally, ALPHA wants to improve service quality and process time throughout the entire value chain. BETA's goal is to leverage emergent technology in order to create a more attractive and ergonomic workplace. Both seek to implement a software product usable on smart glasses in order to evaluate the innovativeness of the technology in logistics.

Throughout the study, various roles and stakeholders participated. The main activities included researchers from the field of information systems and logistics, domain experts from the two logistics case companies, and software development experts from a software provider focused on logistics solutions. The focus groups were moderated and documented by the participating researchers. Additionally, further domain experts from both case companies were included in an online questionnaire.

3.2 Data Collection

The data collection of this study uses different design techniques and hence, benefits from interaction research design practices [38]. Throughout the project duration, data from multiple sources of evidence was collected, i.e. *expert interviews*, *on-site* *observations through shadowing, focus group sessions, discussions, and an online survey*. An overview of the different design techniques used for data collection during product discovery, and their later use in the analysis is presented in Table 2. The numbers in brackets in the *Data Type* column represents the amount of occurrences (count) of the respective data type. The multiple sources of evidence allow us to triangulate results. Mostly, the literature review and ten discussions in the setting of focus groups lead to the definition of potential system components, collected as idea proposals and vision elements. The collection of data follows the guidelines set by Yin [35], i.e. using multiple sources of evidence, creating a case study database, and maintaining a chain of evidence.

Table 2. Overview of design techniques and their use in data analysis.

Design Techniques	*Data Type (Count)*	*Use in Analysis*
Expert interviews	Interviews with logistics IS experts (2)	- Gather initial ideas for the application of smart glasses in logistics. - Investigate user as an information source.
Focus groups / shadowing	Warehouse observations (4 *together with idea proposal discussions*)	- Examine and gain insights into the usage context.
	Idea proposal discussions (4 *together with warehouse observations*)	- Increase understanding of further potential idea proposals for smart glasses applications. - Investigate collaborative approaches through group discussion.
	Idea cost-benefit discussions (1)	- Construct and analyze cost-benefit based on ideas and product visions.
	Idea proposal consolidation (1)	- Corroborate understanding and consolidate analyzed idea proposals.
	Product vision formulation (1)	- Corroborate understanding and enrich idea proposals with further information and examples.
	Requirements formulation (3)	- Evaluation and formulation of the previously defined ideas and vision elements as requirements from both technical and functional perspectives.
Online questionnaire	Vision survey (1)	- Corroborate product vision benefits with other stakeholders. - Prioritize product vision elements and requirements.
Practical development and design	Design Mockups (3)	- Integrate and crosscheck understanding and findings using design artefacts. - Investigate participatory approaches to user participation.

Design Techniques	*Data Type (Count)*	*Use in Analysis*
	Low-fidelity prototypes (3)	- Integrate feedback and corroborate understanding of design. - Integrate and crosscheck interaction scenarios.
Prototype Evaluation	Focus group discussions (2) Survey (1)	- Review project instantiation through user feedback. - Gain information and feedback on the usage context and acceptability.

The following paragraphs provide a brief description of the applied design techniques embedded in the respective software discovery phases.

Propose Initial Ideas. Initial ideas were identified by conducting expert interviews [39]. As part of the *initialization*, two experts were interviewed, a business system consultant and a business system analyst with experience in logistics. The interviews were each scheduled for two hours and were structured by previously developed guidelines. The interviews helped to understand logistics processes, the respective technical requirements to the technology of smart glasses, and practical development requirements in the domain of smart glasses. Additional usage scenarios were then collected by shadowing employees of both companies according to Myers [40]. Set up as a three-day-observation of workflows and activities relevant for the handling of cargo in goods receipt, storage picking and goods issue, results were documented as field notes and process models.

User Engagement. The *user engagement* phase was started through different focus groups. Eight focus groups were held according to established guidelines [41]. Employees of both companies, researchers with expertise in smart glasses technology and in logistics, and representatives of a software company were present at all focus group meetings. Each focus group was conducted moderated and extensively documented by a team member of the research team. A meeting protocol was sent to all participants, allowing them to make corrections and adjustments to the notes.

Product Vision Building. In order to build the respective *product vision* incorporating the different system components, a catalogue of 36 idea proposals was developed and enriched by descriptions and examples to generate common understanding. These particularized ideas formed vision elements, which were then evaluated using a personalized online survey. In addition to general demographic data and entry-fields for comments and notes, the questionnaire consisted of four questions for each vision (general usefulness, innovativeness, personal usefulness, adoption) using a 7-point Likert scale. 31 participants completed the survey (67% domain experts, 7% software developers, 26% researchers).

Requirements Specification. In order to *specify requirements*, the evaluation results were combined with an additional technical focus group meeting, discussing technical requirements of each vision element. Stemming from this inductive combination, a prioritization of the requirements was conducted in a backlog. Resulting from further focus group discussions with domain experts and software developers, functional and technical requirements were formulated, both for general applicability for smart glasses-based software, and for specific guidance of the respective vision element instantiation.

Up-Front Development and Design. Through the *implementation and design* of three different prototypes, the respective vision elements were instantiated iteratively. Three prototypes were implemented with two different goals aiming at supporting employees from logistics at their daily tasks through different sets of functions (documentation of damages or assembly errors, and process visualization for value-added services). The latter prototype was implemented in two iterations. The implemented systems were evaluated by applying different methods for each system. While the initial prototype was evaluated through a survey with students, the other two instantiations were presented and discussed with domain experts. The various qualitative remarks made by the participants as part of subsequent iterations of the phases *user engagement*, *product vision building,* and *requirements specification* were then included in the next iteration of development and design.

3.3 Data Analysis

The analysis was initiated by applying inductive class formation according to Mayring [42] for each data set, generating idea proposals for potential smart glasses application scenarios on the domain of logistics. Ideas were gathered investigating (a) functional support of a smart glasses application, (b) an implementation of technical requirements in logistics, or (c) a potential or already proven application scenario. Suitable ideas were identified in a subsequent analysis, whereas ideas could be assigned to an already defined idea (subsumption). In case no equivalent idea was previously proposed, a new element was formed inductively based on the specific content [40]. A fundamental challenge in the continuous formulation and aggregation of the idea proposals was the occurrence of partly diverging degrees of abstraction in the description through the different methods. Hence, category formation according to Mayring [42] was applied again, based on the results of the different methodological approaches, in order to subsume the idea proposals gradually. Thereafter, descriptions were derived (including actors, processes, and activities), examples for all scenarios were developed by domain experts of the two case companies, and product vision elements were generated. These artefacts were validated and consolidated in additional focus group meetings and formed the basis for the definition of technical requirements. The line of argumentation for the technical assessment of the requirements was inductively formed. The purpose of this coding step was to consolidate and derive the technical design and create architectural design proposals [42]. Pattern matching and explanation building was applied in order to analyze and synthesize our findings.

4 Findings from a Software Discovery Project

As we seek to investigate the role of user participation during the discovery of software products for smart glasses, the results are presented along the different types of user participation. First, we present findings supporting the participation of users as an information source. Second, we report findings with the user as a co-creator. Third, we present findings concerning the user as an innovator.

4.1 The User as Source of Information

Users were involved as a key source of information through the execution of interviews. This was the first design technique we applied during data collection at an early stage of software discovery. Following the recommendation of gathering idea proposals in software discovery, we identified first ideas, which were then further evaluated. We documented and collected all idea proposals in an idea catalogue. Hence, the idea generation reflects the initialization phase of software discovery, with the catalogue being equivalent to the idea pool. As a result of shadowing and observing users, they provided us with additional insights and information (e.g. at which process steps the user needs both hands). An idea proposal stemming from these activities was a system to support the picking process, by displaying information on which product to pick next, and provide confirmations via barcode scanning.

4.2 The User as Co-Creator

Thereafter, warehouse tours helped to gain more detailed insights and impressions of the needs and processes of different warehouse settings. After two tours at each customer company, followed by in-depth discussions with various stakeholders, most of the previously gathered ideas were validated. Furthermore, new ideas were identified and discussed. The results from our initial user participation activities (i.e., the expert interviews, shadowing, and initial focus groups), formed the first version of our idea pool. Since the focus groups included stakeholders and potential users, the tours and discussions engaged the users as part of the software discovery process.

Throughout the participatory discussion and idea pool evaluation, various ideas were regarded as either too ambiguously defined, or too similar to other ideas. To promote a common understanding across both companies, each company was asked to enrich each element with a detailed description and possible scenarios. This helped us to identify differences between both case companies. As a result, when larger differences were identified, we divided an idea into separate elements (e.g., displaying warnings and safety instructions was separated to a process-based and an object-based warning system idea). Thereafter, we could unify many elements that covered the same area, yet were initially associated with two different idea proposals (e.g., the ideas of identifying objects through QR- and barcodes or RFID were unified to a general object identification). Through this consolidation, the final idea pool was reached for further investigation. The underlying warehouse tours and the focus group discussions were planned in the first iteration and served as main basis for all subsequent steps. Detailed evaluation of the resulting ideas and vision elements was conducted toward the middle and end of the discovery phase. From this type of user participation, more elaborate ideas and vision elements could be derived, such as the documentation of damages by usage of the smart glasses' camera and the transfer of this documentary evidence to damage management. These elements form the product backlog that guided the development, and were later instantiated as a prototype.

4.3 The User as Innovative Force behind Trends

With a well-described and communicated set of system components, a prioritization according to their practical benefit was needed. Hence, an online questionnaire was developed to provide a second evaluation. For both companies, the survey led to a prioritized list of the previously unsorted backlog items. Through this survey, users provided their input and evaluated each component. They were given the means to engage and prioritize backlog items according to the stakeholders' needs. This enabled the user to actively shape the software discovery process, as backlog elements that were deemed as unimportant or not useful for the customer and user could be neglected. This increased the influence of the user, in particular in the area of emergent technologies. The survey counts toward requirements specification, as specific requirements or concerns of different stakeholders and users were highlighted.

Mock-ups as well as prototypes were then designed, developed and evaluated in order to showcase the application of smart glasses in a specific use case (process guidance). Benefitting from early user and stakeholder participation, we were able to document cases in which idea proposals, their detailed formulation as vision elements and development and design artifacts were created by stakeholders or users. For example, the idea to support the process of assembling, equipping and examining promotional displays for groceries was actively developed and matured by a stakeholder and participant of the warehouse tours, and was then chosen as first content for the process guidance prototype. This prototype was then instantiated in two iterations. During both iterations, a close collaboration with the users existed, leading to two maturity stages. The development was conducted toward the end of a product discovery iteration.

5 Toward the Process Model for User Participation

We developed a preliminary process model that seeks to clarify the interrelations of different types of user participation (Figure 2). Through the previously described activities, users and stakeholders were actively involved in the discovery process. Our results suggest that early participation of users and stakeholders is exceedingly important in emergent technologies, as desirability is essential for an intention to use a new system.

The process model starts with the intention to develop a new product idea, serve a new user need, or utilize a new technology. Hence, the development and design team has to learn more about the users' tasks and work context. In this scenario, the users typically act as a source of information for the development team. Building on this knowledge, the users can be invited to co-create, for example through co-creation workshops. Through co-creation design techniques, the users becomes more familiar with essentials of the software development and design processes. Once familiar with different techniques and an understanding of their individual strengths and weaknesses, the users may further contribute. They are now equipped with essential process skills and corresponding tools, which are typically readily available online for easy and quick sketches and prototypes. Building on these skills, the users can develop their own design suggestions and share those with the development and design team. Also, new feature ideas can be integrated into existing screenshots by the users, simplifying the communication between the users and the development team. Therefore, the users can drive further innovation by developing their own ideas. We suggest that the evolution of different user participation types is the result of a knowledge creation process [43], where knowledge about the application domain and knowledge about the technology need to be combined [44]. Through structuring this approach as a sequential and incremental process model with iterations between the subsequent stages (i.e. having one stage of user participation leading to a more mature, more integrated stage including changes to the user's tasks, goals and the environment), we provide an ex-post suggestion on how a software discovery project can be conducted. While the definition of a chronological order of the stages is only one possibility of a respective process model, it provided us with an easy to use and successful guideline throughout the investigated project.

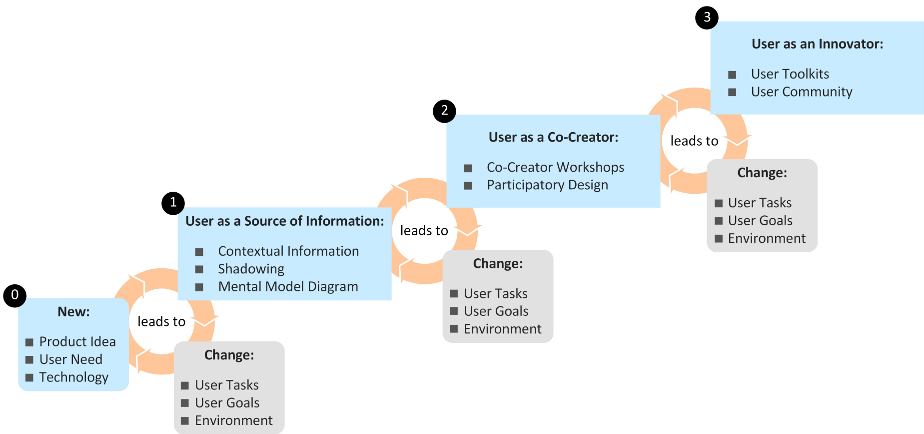


Figure 2. Preliminary process model for the evolution of user participation types

While other researchers, such as Damodaran [2] or Markus and Mao [15], stress the importance of user participation and present detailed role descriptions and research propositions, they do not specify any order or evolution of different types of user participation. As our model focuses on the creation of completely new software in the domain of emergent technology, this temporal link serves as a basis for a holistic approach to user participation. The users are thoroughly guided toward more autonomous roles. We observed that without a chronological order or procedural guidelines, users are quickly overwhelmed by the sheer number of design techniques and design dimensions available and lose track of key objectives, leading to scope creep with ideas and functionality that is out of the scope of the system under development.

6 Lessons Learned

We answer our research question with three lessons learned we identified through the course of the case study. Adhering to and considering these three aspects, practitioners can facilitate user participation when discovering new software.

Consider Evolving Types of User Participation. During the study, multiple iterations of software discovery have been conducted. Through the different iterations, different types of user participation have been observed [13], [16]. The data collection started with interviews, which provided a better understanding of the users' tasks and their environment. Within these interviews, the user served as an information source. Thereafter, on-site observations and focus group discussions were the main source of information. Hence, the user's role evolved, as they were not only the source of information, but also participated in the creation and shaping of designs and prototypes. Lastly, users were engaged as they were given the opportunity to actively influence the design of the software, by generating their own design suggestions. Practitioners should keep these different phases in mind when integrating users to gain the desired information or result.

Enhance Desirability through User Participation. We were able to identify early user and stakeholder participation to be essential for a successful discovery of smart glasses software. The early participation helped to shed light on some concerns and other issues raised by the users and stakeholders. Consequently, these issues and concerns could be addressed from the beginning. Involving users early in the process provides assurance and reduces uncertainty for all people involved [17]. For example, many of the idea proposals resulted from the focus group discussions. Several ideas were related to a tension field (e.g., data privacy). Hence, users had to be reassured that the project aimed at improving processes and ergonomics in their favor without jeopardizing their data privacy. Thus, rushed introductions of technological advancements can have negative effects, whereas comprehensive up-front communication and preparation help to avoid such pitfalls. In the case of emergent technologies, uncertainty or unawareness frequently have to be managed. As different stakeholders, such as users or customers, were not able to imagine the specific use of smart glasses in their work context, hands-on demonstrations were provided in order to specify the technological capabilities of smart glasses within their work environment. Through this course of action in conjunction with the innovative characteristics of smart glasses, different desirable prototypical implementations have been developed. This became apparent as new participants could be regularly included into the focus groups.

Thoroughly Discover Software Products for Emergent Technologies. Through the interaction with users and other stakeholders, prior concerns with the technology became apparent. We wanted to put an emphasis on discovery activities using prototypes and mockups. These artefact-centered means of communication helped to reduce uncertainty by specifying design ideas and concepts. While prior research suggests steps and activities for the discovery of products (e.g. [1], [10]), the case context suggests specific nuances for the application with smart glasses. Given that smart glasses are still an emergent technology, little application examples exist. The role of very early participation of users in order to understand the tasks, goals and work context are important in contrast to application domains with a variety of mature software alternatives fulfilling the same or a similar purpose. Hence, while in an established application domain given instances may swiftly consider co-creation workshops or user participation with a focus on innovative ideas, in our case, technology provider and technology user have to become familiar with each other's domains, as users from all stakeholders issued not being familiar with the technology.

7 Conclusion

The paper investigates the discovery of software products for smart glasses. Key phases and activities for software discovery applied in a case study in the logistics context. As a result, the use of different user participation techniques, depending on the products maturity is suggested. While initially, the user is often seen as a source of information for the developers to understand the task goals and work environment. Subsequently, the user can be engaged as a co-creator. With the use of toolkits, the user can also drive innovative processes by developing own design alternatives and ideas for new product features. While we investigated different sources of information over a longer period of time in order to develop a preliminary process model, the study does not investigate cause-and-effect relations. Hence, future research may investigate the use of different user participation types toward the performance of the discovered product.

References

1. Werder, K., Zobel, B., Maedche, A.: PDISC – Towards a Method for Software Product DISCovery. In: Maglyas, A. and Lamprecht, A.-L. (eds.) International Conference on Software Business. pp. 47–62. Springer International Publishing, Cham, CH (2016)
2. Damodaran, L.: User involvement in the systems design process - a practical guide for users. Behav. Inf. Technol. 15, 363–377 (1996)
3. Heikkilä, V.T., Paasivaara, M., Rautiainen, K., Lassenius, C., Toivola, T., Järvinen, J.: Operational release planning in large-scale scrum with multiple stakeholders - A longitudinal case study at F-secure corporation. Inf. Softw. Technol. 57, 116–140 (2015)
4. Kuusinen, K.: Improving UX Work in Scrum Development: A Three Year Follow-Up Study in a Company. In: 5th IFIP WG 13.2 International Conference, HCSE 2014. pp. 259–266. Springer Berlin Heidelberg, Paderborn, DE (2014)
5. Brhel, M., Meth, H., Maedche, A., Werder, K.: Exploring principles of user-centered agile software development: A literature review. Inf. Softw. Technol. 61, 163–181 (2015)
6. Salah, D., Paige, R.F., Cairns, P.: A Systematic Literature Review for Agile Development Processes and User Centred Design Integration. Proc. 18th Int. Conf. Eval. Assess. Softw. Eng. (EASE '14), London, United Kingdom, 13-14 May, 2014. 5:1--5:10 (2014)
7. Bosch, J.: Speed, Data, and Ecosystems: The Future of Software Engineering. IEEE Softw. 33, 82–88 (2016)
8. Nwaka, S., Hudson, A.: Innovative lead discovery strategies for tropical diseases. Nat. Rev. Drug Discov. 5, 941–955 (2006)
9. Cooper, R.G.: Winning at New Products: Creating Value Through Innovation. Basic Books, New York, NY, USA (2011)

10. Stevens, E.: Fuzzy front-end learning strategies: Exploration of a high-tech company. Technovation. 34, 431–440 (2014)
11. Cagan, M.: Product Discovery, http://www.svpg.com/product-discovery
12. Kaulio, M. a.: Customer, consumer and user involvement in product development: A framework and a review of selected methods. Total Qual. Manag. 9, 141–149 (1998)
13. Werder, K.: Team agility and team performance - The moderating effect of user involvement. 24th Eur. Conf. Inf. Syst. ECIS 2016. (2016)
14. Simonsen, J., Robertson, T.: Routledge International Handbook of Participatory Design. Routledge, New York, NY (2013)
15. Markus, M.L., Mao, J.-Y.: Participation in Development and Implementation - Updating an Old, Tired Concept for Today's IS. J. Assoc. Inf. Syst. 5, 514–544 (2004)
16. Cui, A.S., Wu, F.: Utilizing customer knowledge in innovation: antecedents and impact of customer involvement on new product performance. J. Acad. Mark. Sci. (2015)
17. Bano, M., Zowghi, D.: A systematic review on the relationship between user involvement and system success. Inf. Softw. Technol. (2014)
18. Liu, X., Werder, K., Maedche, A., Mädche, A.: A Taxonomy of Digital Service Design Techniques. In: 2016 International Conference on Information Systems, ICIS 2016. pp. 1–12. AISeL, Dublin, IE (2016)
19. Matthies, D.J.C., Haescher, M., Alm, R., Urban, B.: Properties Of A Peripheral Head-Mounted Display (PHMD) Peripheral Head-Mounted Display (PHMD). 1–6 (2015)
20. Frick, W.: Google's Strategy vs. Glass's Potential.
21. McKinsey Digital: Industry 4.0 - How to navigate digitization of the manufacturing sector. (2015)
22. Niemöller, C., Metzger, D., Fellmann, M., Özcan, D., Thomas, O.: Shaping the Future of Mobile Service Support Systems - Ex-Ante Evaluation of Smart Glasses in Technical Customer Service Processes. In: Mayr, H.C. and Pinzger, M. (eds.) Informatik von Menschen für Menschen (Informatik 2016). pp. 753–767. LNI 259, Klagenfurt, Österreich (2016)
23. Lewis, J.R.: Usability: Lessons Learned … and Yet to Be Learned. Int. J. Hum. Comput. Interact. 30, 663–684 (2014)
24. Ralph, P.: Toward Methodological Guidelines for Process Theories and Taxonomies in Software Engineering. IEEE Trans. Softw. Eng. (2018)
25. Pavlou, P. a., El Sawy, O. a.: From IT Leveraging Competence to Competitive Advantage in Turbulent Environments: The Case of New Product Development. Inf. Syst. Res. 17, 198–227 (2006)
26. Mintzberg, H.: Patterns in strategy formation. Manage. Sci. 24, 934–948 (1978)
27. Kajko-Mattsson, M., Nyfjord, J.: A model of agile evolution and maintenance process. Proc. 42nd Annu. Hawaii Int. Conf. Syst. Sci. HICSS. 1–10 (2009)
28. Sy, D.: Adapting Usability Investigations for Agile User-Centered Design. J. Usability Stud. 2, 112–132 (2007)
29. Salah, D., Paige, R., Cairns, P.: A Practitioner Perspective on Integrating Agile and User Centred Design. Proc. 28th Int. BCS Hum. Comput. Interact. Conf. HCI 2014 - Sand, Sea Sky - Holiday HCI. 100–109 (2014)
30. Barki, H., Hartwick, J.: Rethinking the concept of user involvement. MIS Q. 13, 53–63 (1989)
31. Harris, M., Weistroffer, H.: A New Look at the Relationship between User Involvement in Systems Development and System Success. Commun. Assoc. Inf. Syst. 24, 739–756 (2009)
32. Lin, W.T., Shao, B.B.M.: The relationship between user participation and system success: a simultaneous contingency approach. Inf. Manag. 37, 283–295 (2000).

33. McKeen, J., Guimaraes, T., Wetherbe, J.: The relationship between user participation and user satisfaction: an investigation of four contingency factors. MIS Q. 18, 427–451 (1994)
34. Tait, P., Vessey, I.: The effect of user involvement on system success: a contingency approach. MIS Q. 12, 91–108 (1988)
35. Yin, R.K.: Case Study Research: Design and Methods. Sage, Thousand Oaks (2009)
36. Reif, R., Günthner, W.A., Schwerdtfeger, B., Klinker, G.: Pick-by-Vision Comes on Age: Evaluation of an Augmented Reality Supported Picking System in a Real Storage Environment. In: Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa. pp. 23–31. , New York, NY, USA (2009)
37. Dubé, L., Paré, G.: Rigor in information systems positivist case research: current practices, trends, and recommendations. MIS Q. 27, 597–635 (2003)
38. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07. p. 493. ACM Press, New York, New York, USA (2007)
39. Bogner, A., Menz, W.: The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction BT - Interviewing Experts. Presented at the (2009)
40. Myers, M.: Qualitative Research in Business & Management. Sage Publications Ltd., London (2009)
41. Bohnsack, R.: Group discussion and focus groups. A companion to Qual. Res. 210–221 (2004)
42. Mayring, P.: Qualitative Content Analysis: Theoretical Foundations, Basic Procedures and Software Solution. (2014)
43. Joshi, A.W., Sharma, S.: Customer Knowledge Development: Antecedents and Impact on New Product Performance. J. Mark. 68, 47–59 (2004)
44. von Hippel, E., Katz, R.: Shifting Innovation to Users via Toolkits. Manage. Sci. 48, 821–833 (2002)

The Fluidity of the Self-Concept as a Framework to Explain the Motivation to Play Video Games

Bastian Kordyaka[1], Marius Müller[1], and Björn Niehaves[1]

[1] University of Siegen, Chair of Information Systems, Siegen, Germany
(Bastian.Kordyaka, Marius.Mueller, Bjoern.Niehaves)@uni-siegen.de

Abstract. A better understanding of the motivation to play video games and potential antecedents have a long history in Human Computer Interaction research. Besides different motivational dimensions specific to video games, researchers already used the personality of players to explain the motivation to play and the subsequent video game use. At this juncture, they postulated a rather static self-concept underlying the personality of players. The study at hand tries to resolve this shortcoming and proposes a more holistic perspective on personality following the assumptions of the Social Identity Approach from psychology, which postulates a much more fluid and context-specific salient parts of the self-concept. Specifically, we use findings from consumer research arguing that the dimensional fit between the perception of the self-concept of a player and the corresponding video game holds the potential to explain the motivation to play as well as the subsequent usage of the video game.

Keywords: Video Games, Motivation, Personality Traits, Self-Concept.

1 Introduction

In this day and age, video games can be considered as a specific and especially popular form of socio technical systems [1]. Within the last decade, they experienced an upturn regarding their popularity. In 2017 more than 2.2 billion people worldwide played video games and the industry had an estimated global revenue of $108.9 billion [2]. This indicates that a better understanding of the hedonic motivation to play video games and the subsequent use is a crucial question for academic research and practice.

Information Systems (IS) and Human-Computer Interaction (HCI) research already captured hedonic motivation as a driver for the use of sociotechnical systems (e.g. live streaming, ecommerce) [3, 4]. Additionally, research already explored motivational drivers to play video games. Pioneering work in this regard, identified different clusters of relevant motivations in the context of video games and looked for predictors of motivation [5]. Notable examples used the personality of players as potential predictors of motivation. Surprisingly, the majority of empirical studies postulated a rather static self-concept of personality including demographic or context unspecific personal traits, which limits the explanatory power significantly [5–8].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

With our study, we aim to address this shortcoming and follow the assumptions of the Social Identity Approach (SIA) from psychology postulating a much more fluid and context dependent self-concept of players [9, 10]. We plan to use existing findings from consumer research and test them for the first time in the context of video games [11–13]. Specifically, we propose the fit between the player's self-concept in the specific domain of a video game as a relevant antecedent for the motivation to play and the subsequent use. Providing a more holistic and relational explanation for video game use promises several important contributions. First, it allows researchers to better understand one contemporary especially meaningful form of technology use and transfer the findings to neighboring IS relevant contexts (e.g. health, learning). Second, it provides the gaming industry with the opportunity to learn more about the design of a game and gain insights into hedonic motivation as a driver for economic success. To ensure the external validity of our findings and explore differences and commonalities between different games, we intend to examine the three most successful games of the year 2017 (League of Legends, Fortnite: Battle Royal, and Overwatch) [14]. Additionally, we want to make use of a multi-level analysis using the levels of single games and aggregate the findings on a higher level of all three games. Therefore, the short paper is guided by the following research question:

RQ: *Can the self-concept of a player explain the motivation to play and the subsequent video game usage?*

2 Related Work

2.1 Game Use

In the specific context of video game use, two different streams of research can be detected. First, negative issues like pathological use and addiction [15, 16], violence [17, 18], and physical correlates [19, 20] are oftentimes dealt with. Second, based on the psychology of action [21, 22], several studies proposed different motivational drivers to play video games [8, 23]. We anchor our study within the second stream since we want to expand the current understanding of the motivation to play video games.

2.2 Gaming Motivations

Looking for potential predictors of video game use different motivational models were already proposed using different theoretical underpinnings. Two noteworthy approaches in this regard are the self-determination theory [24, 25] and the uses and gratifications theory (UGT) from media psychology [26, 27].

We build our study around the UGT since we target to theoretically enrich the existing status-quo of video game motivation, which allows for a flexible approach. The UGT is a widely accepted taxonomy of six main motivations (action, social, mastery, achievement, immersion, and creativity) to play video games [5].

2.3 Predictors of Motivation

According to the UGT, players actively seek to satisfy their needs with their behavior [26]. Therefore, the choice of behavior largely depends on the player's personality consisting of personal traits and the self-concept, which can be understood as UGT components predicting motivation and usage [27].

Personality traits. One of the most established models to categorize personality traits is the five-factor model of personality, which has been used as motivational predictors in several game related studies [6, 28]. The Big Five taxonomy assumes a rather static personality of players and consists of the dimensions openness, conscientiousness, extraversion, agreeableness, and neuroticism [29].

Self-concept. The self-concept can be understood as the totality of the individual's thoughts and feelings referencing to himself/herself as an object [30]. Considering the self-concept from a SIA perspective, it can be characterized by its fluid and context-specific salience [9, 10]. Therefore, we assume that game use largely depends on an individual's own perception of how well a certain game is able to satisfy needs and the corresponding self-concept. The following dimensions are commonly used to describe the self-concept: actual self, ideal self, social self, and the ideal social self [11, 13].

2.4 Relevant Contexts

We intend to examine the three most successful games of the year 2017 *League of Legends*, *Fortnite: Battle Royal*, and *Overwatch*. For comprehensive overviews of the games, we refer to previous literature providing detailed descriptions [31–33].

3 Methodological Approach

3.1 Research Design

We plan on using a cross-sectional approach. Therefore, we will use an online survey to collect self-reported data and covariance based statistics. Additionally, we will make use of a pre-study to develop a measurement (semantic differential) of relevant dimensions related to the games and to have a benchmark to compare them to the different parts of the self-concept. Figure 1 shows our research model and hypotheses.

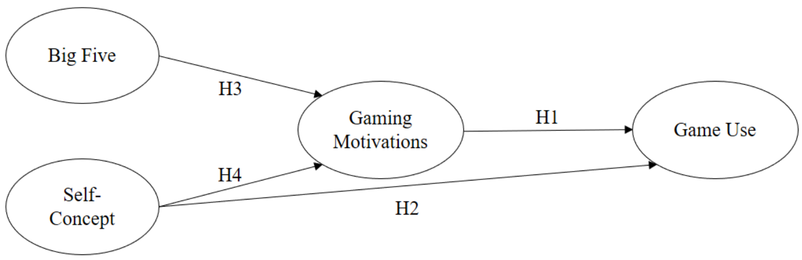


Figure 1. Research Model

3.2 Data Collection

In order to ensure conclusive results, we will survey players of the three games *League of Legends*, *Fortnite: Battle Royal*, and *Overwatch*. We want to consult ordinary players, since the aim of our study is about stereotypical use patterns. In order to acquire a significant amount of respondents (we plan to acquire at least > 200), we will use different channels (e.g. community boards, social media, gatekeepers) to disseminate the link to our survey and promise different forms of incentives (e.g. game currency vouchers) to ensure the motivation of participants in our study.

3.3 Data Analysis

To analyze the data, we aim to make use of different statistical tools. First, we will use co-variance based path modeling to test the hypotheses. Second, we intend to carry out a multilevel analysis to compare findings from the lower level of single games and aggregate them to a higher level.

3.4 Measurements

To measure the variables of our study, we will adapt empirically validated scales to the context of our study (see Table 1). Additionally, we will measure demographic and control variables to have the chance to control our results for potential confounds.

Table 1. Measurements of the study

Type of variable	*Name of variable*	*Exemplary wording of item*	*Source*
Dependent	Game Use	Please indicate your frequency of play? (5 items)	[34]
Mediating	Gaming Motivations	I like chatting with others (12 items).	[35]
Independent	Big Five	I am someone who is full of energy (30 items).	[36]
	Self-Concept	I have respect for myself (35 items).	[37]

4 Outlook

The paper at hand proposes a holistic approach to capture the motivation to play and the subsequent use of different video games. This leads to different implications. First, capturing the effects on different levels of different games indicates the potential for additional insights (e.g. commonalities and differences between games). Second, illustrating the meaningfulness of the self-concept promises fruitful avenues for theory (e.g. a context specific theory to explain video game use) and practice (e.g. demand and economic meaning). Besides the significant insights, the proposed study includes several limitations. First, we will not have the chance to identify causal connections between the constructs because we plan a survey. Nonetheless, on the basis of our results it is possible to conduct experiments to test the causality of relationships. Second, the amount of items is rather high (> 100) which might limit the response rates. To avoid this shortcoming, we plan to use different types of incentives.

References

1. Detering, S.: Eudaimonic Design, or: Six Invitations to Rethink Gamification. 27.
2. Newzoo: The Global Games Market 2017 | Per Region & Segment, https://newzoo.com/insights/articles/the-global-games-market-will-reach-108-9-billion-in-2017-with-mobile-taking-42/.
3. Cai, J., Wohn, D.Y., Mittal, A., Sureshbabu, D.: Utilitarian and Hedonic Motivations for Live Streaming Shopping. In: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '18. pp. 81–88. ACM Press, SEOUL, Republic of Korea (2018).
4. Venkatesh, V., Thong, J.Y., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS quarterly. pp. 157–178 (2012).
5. Yee, N.: The Gamer Motivation Profile: What we Learned from 250,000 Gamers. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. pp. 2–2. ACM (2016).
6. Johnson, D., Gardner, J.: Personality, Motivation and Video Games. In: Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction - OZCHI '10. p. 276. ACM Press, Brisbane, Australia (2010).
7. Jeng, S.-P., Teng, C.-I.: Personality and Motivations for Playing Online Games. Social Behavior and Personality: An International Journal. 36, 1053–1060 (2008).
8. Yee, N.: Motivations for play in online games. CyberPsychology & behavior. 9, pp. 772–775 (2006).
9. Tajfel, H., Turner, J.C.: The Social Identity Theory of Intergroup Behavior. (2004).
10. Baumeister, R.F., Hutton, D.G.: Self-Presentation Theory: Self-Construction and Audience Pleasing. In: Mullen, B. and Goethals, G.R. (eds.) Theories of Group Behavior. pp. 71–87. Springer New York, New York, NY (1987).
11. Achouri, M., Bouslama, N.: The Effect of the Congruence between Brand Personality and Self-Image on Consumer's Satisfaction and Loyalty: A Conceptual Framework. IBIMA Business Review Journal. pp. 1–17 (2010).
12. Souiden, N., M'Saad, B., Pons, F.: A Cross-Cultural Analysis of Consumers' Conspicuous Consumption of Branded Fashion Accessories. Journal of International Consumer Marketing. 23, pp. 329–343 (2011).
13. Jamal, A., Goode, M.M.H.: Consumers and Brands: A Study of the Impact of Self-Image Congruence on Brand Preference and Satisfaction. Marketing Intelligence & Planning. 19, pp. 482–492 (2001).
14. The Most Popular PC Games Right Now, //www.ranker.com/list/most-popular-pc-games-today/ranker-games.
15. Gentile, D.A., Choo, H., Liau, A., Sim, T., Li, D., Fung, D., Khoo, A.: Pathological Video Game Use Among Youths: A Two-Year Longitudinal Study. PEDIATRICS. 127, e319–e329 (2011).
16. Sim, T., Gentile, D.A., Bricolo, F., Serpelloni, G., Gulamoydeen, F.: A Conceptual Review of Research on the Pathological Use of Computers, Video Games, and the Internet. International Journal of Mental Health and Addiction. 10, pp. 748–769 (2012).
17. Olson, C.K., Kutner, L.A., Warner, D.E., Almerigi, J.B., Baer, L., Nicholi, A.M., Beresin, E.V.: Factors Correlated with Violent Video Game Use by Adolescent Boys and Girls. Journal of Adolescent Health. 41, pp. 77–83 (2007).

18. Ferguson, C.J., Olson, C.K.: Video Game Violence Use Among “Vulnerable” Populations: The Impact of Violent Games on Delinquency and Bullying Among Children with Clinically Elevated Depression or Attention Deficit Symptoms. Journal of Youth and Adolescence. 43, pp. 127–136 (2014).
19. Vandewater, E.A., Shim, M., Caplovitz, A.G.: Linking Obesity and Activity Level with Children’s Television and Video Game Use. Journal of Adolescence. 27, pp. 71–85 (2004).
20. Mentzoni, R.A., Brunborg, G.S., Molde, H., Myrseth, H., Hetland, J., Pallesen, S.: Problematic Video Game Use: Prevalence and Associations with Mental and Physical Health. Cyberpsychology, Behavior, and Social Networking. 14, pp. 591–596 (2011).
21. Heckhausen, H.: Achievement motivation and its constructs: A cognitive model. Motivation and emotion. 1, pp. 283–329 (1977).
22. Gollwitzer, P.M., Bargh, J.A.: The Psychology of Action: Linking Cognition and Motivation to Behavior. Guilford Press (1996).
23. Przybylski, A.K., Rigby, C.S., Ryan, R.M.: A Motivational Model of Video Game Engagement. Review of General Psychology. 14, pp. 154–166 (2010).
24. Ryan, R.M., Deci, E.L.: Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. American Psychologist. 55, 68 (2000).
25. Ryan, R.M., Rigby, C.S., Przybylski, A.: The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motivation and Emotion. 30, pp. 344–360 (2006).
26. Sherry, J.L., Greenberg, B.S., Lucas, K., Lachlan, K.: Video Game Uses and Gratifications as Predictors of Use and Game Preference. 13 (2006).
27. Ruggiero, T.E.: Uses and Gratifications Theory in the 21st Century. Mass Communication and Society. 3, pp. 3–37 (2000).
28. Park, S., Chung, N.: Mediating Roles of Self-Presentation Desire in Online Game Community Commitment and Trust Behavior of Massive Multiplayer Online Role-Playing Games. Computers in Human Behavior. 27, pp. 2372–2379 (2011).
29. Barrick, M.R., Mount, M.K.: The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. Personnel psychology. 44, pp. 1–26 (1991).
30. Rosenberg, M., Schooler, C., Schoenbach, C., Rosenberg, F.: Global Self-Esteem and Specific Self-Esteem: Different Concepts, Different Outcomes. American Sociological Review. 60, pp. 141–156 (1995).
31. Ferrari, S.: From Generative to Conventional Play: MOBA and League of Legends. In: DiGRA Conference (2013).
32. Blizzard: Overwatch, https://playoverwatch.com/de-de/.
33. Epic: Hol dir Fortnite, https://www.epicgames.com/fortnite/de/buy-now/battle-royale.
34. Taylor, S., Todd, P.A.: Understanding Information Technology Usage: A Test of Competing Models. Information Systems Research. 6, pp. 144–176 (1995).
35. Yee, N., Ducheneaut, N., Nelson, L.: Online Gaming Motivations Scale: Development and Validation. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 2803–2806. ACM (2012).
36. Pervin, L.A., John, O.P.: Handbook of Personality: Theory and Research. Elsevier (1999).
37. Hattie, J.: Self-Concept. Psychology Press (2014).

Heart over Heels? An Empirical Analysis of the Relationship between Emotions and Review Helpfulness for Experience and Credence Goods

Michelle Müller[1], Dominik Gutt[1]

[1] Paderborn University, Faculty of Business Administration and Economics, Paderborn, Germany
{mmuell15}@mail.uni-paderborn.de, {dominik.gutt}@upb.de

Abstract. According to current scholarly and practitioner thinking, one way of enhancing the perceived helpfulness of reviews is by encouraging the use of emotional language. Yet, studies on review helpfulness have paid little attention to studying this effect as it applies to different product types, namely experience and credence goods. Using data from amazon.com, we conduct an empirical test using a natural language understanding algorithm. Our results suggest that for both experience and credence goods, fear, joy, and sadness are correlated with an increase in review helpfulness, whereas anger is negatively correlated with it. These emotions are perceived as more helpful for experience goods than for credence goods.

Keywords: Helpfulness, Emotions, Experience Attributes, Credence Attributes

1 Introduction

Human-Computer Interaction (HCI) between online shoppers and online review systems has led to the creation of massive amounts of data. Due to the rapidly growing volume of product reviews on review websites, customers often face information overload so that finding particularly helpful reviews can be like searching for a needle in a haystack [1]. To alleviate this problem, online review systems attempt to identify reviews that customers perceive as most helpful [2]. Consequently, research has investigated the features of helpful reviews [3-5], such as emotions conveyed by review texts. One way of eliciting this knowledge is to nudge reviewers towards using particularly helpful emotional language by presenting them with review templates with suggestions on how to craft a helpful review.

Research suggests that emotions expressed in reviews significantly influence perceptions of helpfulness [5, 6]. Emotions are a common and important reason for electronic word of mouth because consumers often write reviews when they are either extremely satisfied or extremely dissatisfied [7]. For instance, if a reviewer expresses lots of joy about a film, this is perceived as very helpful [5], whereas reviews of products like printers are perceived as more helpful when written in a more neutral tone [6]. Existing research on the topic focuses on products that have either search or

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

experience attributes [5, 6]. However, to the best of our knowledge, research has not yet examined the difference between the impact that emotions have on review helpfulness for goods with experience attributes (those that can more easily be assessed after purchase) vs. goods with credence attributes (which are difficult to assess even after purchase). In this paper, we refer to experience (credence) goods, when a good has a high share of experience (credence) attributes. For both of these product types, reviews seem to be particularly valuable because the degree of uncertainty before purchase is high [5]. Hence, we aim to answer the following research questions:

1. *What impact does the degree of emotionality in online reviews for experience and credence goods have on their perceived helpfulness?*
2. *Does the relationship between the emotional content of a review and its perceived helpfulness differ depending upon the type of goods reviewed?*

Our preliminary results suggest that emotional content in reviews is associated with an increase in perceived helpfulness. Counter to our intuition and second hypothesis, this relationship is more pronounced for experience than for credence goods. Our work thus provides research implications, in particular for HCI, and practical implications for online review system designers who can use our results to improve their review templates by suggesting to reviewers to express discrete emotions in their reviews.

2 Related Literature, Theoretical Background, and Hypotheses

An earlier study has employed a latent semantic analysis (LSA) to identify discrete emotions in reviews [8]. This study finds that emotions like hope, happiness, anxiety, and disgust express certainty and have differential implications for a review's helpfulness. Another study employs a bag of words approach to analyze the association between overall emotional content in online reviews for movies and a review's helpfulness [5], finding a positive relationship between overall emotionality and movie review helpfulness. Finally, another study employs a lexicon- and an expert dictionary-based approach to measure a general relationship between discrete emotions and review helpfulness [9], presenting evidence that emotions are conducive to review helpfulness. However, they do not investigate differences in this relationship across product types.

Our study aims to disentangle the relationship between emotions and experience and credence attributes. Experience attributes can only be inspected after purchase, so the consumer must first have bought the product or service, such as a restaurant meal [10]. Credence attributes are characterized by an even higher level of uncertainty, such that the attribute of the product or service can hardly be assessed even after purchase (e.g., the effectiveness of cosmetics or of a medical treatment) [11]. However, reviews of goods whose assessment involves a significant element of either experience or credence attributes can be especially helpful given the high degree of uncertainty prior to purchase (compared to search goods, for example). Finally, in line with prior literature [9], we choose the discrete emotions of anger, disgust, fear, joy, and sadness as the emotions of interest. Expressing emotions has been demonstrated to be conducive to the helpfulness of reviews [5, 9]. Given the high level of uncertainty for experience and credence goods prior to purchase, we expect to find a positive relationship between emotional review content, especially when the emotions are negative [12], and a review's helpfulness. We hypothesize: *Hypothesis 1: Emotions in online reviews for both experience and credence goods are positively associated with a review's perceived helpfulness.*

After consumption, a customer of a credence good is left with more remaining product uncertainty than a customer of an experience good. Thus, when writing a review, a credence good customer can provide less factual information but arguably has to resort to their feeling about the good after consumption. Therefore, to enhance review helpfulness, reviews for credence goods should provide more emotional content, compared to experience goods. Thus, we hypothesize: *Hypothesis 2: The relationship between emotions and perceived helpfulness of a review is moderated by the product type, with emotions in reviews of credence goods being more helpful than in reviews of experience goods.*

3 Research Setup and Empirical Analysis

We obtained a data set from amazon.com containing 53,218 single online reviews of experience goods (video games, digital cameras, in line with [13, 14]) and of credence goods (health- and fitness apps, health and personal care products, in line with [15, 16]). All the reviews had obtained at least one helpfulness vote. They had been collected in July 2014 and contain all reviews since May 1996. To identify expressed emotions in online reviews, we employ the natural language understanding service of IBM Watson. This service can extract features of unstructured content, such as emotions of online reviews. For each review we obtained data related to anger, disgust, fear, joy, and sadness, as well as an overall emotion score.

To test our hypotheses, we conducted a linear regression model with *HELPFUL* as dependent variable. In line with existing literature [3, 4], *HELPFUL* is measured as the proportion of helpful votes out of the total votes a review has received. Review-related *Control Variables* (vector *X*) comprise the rating (1 to 5), squared rating, the average rating, the number of ratings, word count, review age, and the average readability index. The variable $OVERALL\_EMO$ controls for the overall emotional review sentiment and is defined between -1 (negative sentiment) and +1 (positive sentiment). Formally, estimation equation 1 displays our model for review *i* and product *j*:

$$HELPFUL_{ij} = \beta_0 + \beta_1 ANGER_{ij} + \beta_2 DISGUST_{ij} + \beta_3 FEAR_{ij} + \beta_4 JOY_{ij} + \beta_5 SADNESS_{ij} + \beta_6\, OVERALL\_EMO_{ij} + \gamma X_{ij} + \epsilon_{ij} \tag{1}$$

The coefficients for all dimensions of emotions are consistently larger for experience goods than for credence goods (Table 1). Thus, we find support for H1 (emotions are generally conducive to review helpfulness) but we have to reject H2. We tested the statistical significance of coefficients (*Differences*) using the Chow test and find significant differences for all but one coefficient. For robustness, we tested consumer goods and digital goods separately, implemented product fixed effects, and ran Tobit regressions to account for the truncation of *HELPFUL*. Acknowledging that some studies classify digital cameras as a search product [2], we also ran our regressions again without cameras. Our results remain qualitatively unchanged after all these tests.

Table 1. Linear Regression Results

Model *Variable*	*Experience Goods* *HELPFUL*	*Credence Goods* *HELPFUL*	*Differences* *HELPFUL*
ANGER	—0.05***(0.01)	—0.06 (0.04)	χ^2= 8.96***
DISGUST	0.00 (0.01)	—0.05 (0.04)	χ^2=1.07
FEAR	0.13*** (0.01)	0.08** (0.04)	χ^2=4.47**
JOY	0.14*** (0.01)	0.11** (0.03)	χ^2=25.6***
SADNESS	0.23*** (0.01)	0.08** (0.03)	χ^2=41.95***
OVERALL_EMO	0.09*** (0.01)	0.04*** (0.01)	χ^2=30.30***
Control Variables	✓	✓	-
Observations	46,463	6,531	-
R^2	*0.1739*	*0.0854*	-

Note: Robust standard errors are in parentheses. *p < 0.1; ** p < 0.05; *** p < 0.01.

4 Preliminary Conclusions and Future Research

This study provides preliminary indication that discrete emotions, namely, fear, joy, and sadness, are associated with an increase in review helpfulness for both experience and credence goods. More interestingly, our empirical results also show that these emotions are perceived as more helpful for experience goods than for credence goods, indicating that reviews for products with credence attributes should be written in a more neutral tone, compared to reviews of experience goods. Following the call of prior literature [3-5], review systems can use our approach to improve review templates for their customers by suggesting to reviewers that they should express certain discrete emotions or by discouraging them from using certain other emotions (such as anger or disgust). Ultimately, the preliminary insights from this study provide a first step to improving the design of online review systems. Moreover, this work strengthens current scholarly understanding of how to craft a helpful review [2, 3, 4, 13, 17].

Naturally, as our study is based on observational data, it carries limitations which require extended investigation as follows. First, we plan to further validate the identification of emotions by using another natural language understanding algorithm as well as human coders, since text mining techniques such as IBM Watson should always be used with caution. Second, we plan to conduct additional experiments to rule out confounding factors such as reviewer-review self-selection and factors such as the review reader's perception of the review writer's effort. Third, in order to obtain a more holistic picture, we intend to analyze the impact of emotions for digital and non-digital products, as well as for hedonic and utilitarian goods.

5 Acknowledgements

This work has been partially supported by the German Research Foundation (DFG) within the Collaborative Research Center 901 "On-The-Fly Computing" under the project number 160364472-SFB901.

References

1. Liu, Y., Huang, X., An, A., Yu, X.: Modeling and Predicting the Helpfulness of Online Reviews. Eighth IEEE International Conference on Data Mining, 443-452 (2008)
2. Mudambi, S. M., Schuff, D.: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon. com. MIS Quarterly, 185-200 (2010).
3. Yin, D., Bond, S. D., Zhang, H.: Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews. MIS Quarterly, 38(2), 539-560 (2014)
4. Yin, D., Bond, S. D., Zhang, H.: Keep your Cool or let it out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews. Journal of Marketing Research, 54(3), 447-463 (2017)
5. Ullah, R., Zeb, A., Kim, W.: The Impact of Emotions on the Helpfulness of Movie Reviews. Journal of Applied Research and Technology 13(3), 359-363 (2015)
6. Peng, C.H.,Yin, D.,Wei, C.P., Zhang, H.: How and when Review Length and Emotional Intensity Influence Review Helpfulness: Empirical Evidence from Epinions.Com. Thirty Fifth International Conference on Information Systems (ICIS). (2014)
7. Anderson, E. T., Simester, D. I.: Reviews without a Purchase: Low Ratings, Loyal Customers, and Deception. Journal of Marketing Research 51(3), 249-269 (2014)
8. Ahmad, S.H., Laroche, M.: How do Expressed Emotions Affect the Helpfulness of a Product Review? Evidence from Reviews Using Latent Semantic Analysis. International Journal of Electronic Commerce, 20(1), 76-111 (2015)
9. Felbermayr, A. and Nanopoulos, A.: The Role of Emotions for the Perceived Usefulness in Online Customer Reviews. Journal of Interactive Marketing, 36, 60-76 (2016)
10. Nelson, P.: Information and Consumer Behavior. Journal of Political Economy, 78(2), 311-329 (1970).
11. Darby, M. R., Karni, E.: Free Competition and the Optimal Amount of Fraud. In: The Journal of Law and Economics, 16(1), 67-88 (1973)
12. Sen, S., Lerman, D.: Why are You telling Me this? An Examination into Negative Consumer Reviews on the Web. Journal of Interactive Marketing, 21(4), 76-94 (2007)
13. Huang, P., Lurie, N. H., Mitra, S.: Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods. Journal of Marketing, 73(2), 55-69 (2009)
14. Nelson, P.: Advertising as Information. Journal of Political Economy, 82(4), 729-754 (1974)
15. Dulleck, U., Kerschbamer, R.: On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. Journal of Economic Literature, 44(1), 5-42 (2006)
16. Emons, W.: Credence Goods and Fraudulent Experts. The RAND Journal of Economics, 28(1), 107-119 (1997)
17. Scholz, M., Dorner, V.: The Recipe for the Perfect Review?. Business & Information Systems Engineering, 5(3), 141-151 (2013)

Track 11:

Information Security and Information Privacy

Track Chairs:

Prof. Dr. Ali Sunyaev
Karlsruhe Institute of Technology (KIT)

Prof. Dr. Hanna Krasnova
Universität Potsdam

Unfolding Concerns about Augmented Reality Technologies: A Qualitative Analysis of User Perceptions

David Harborth

Chair of Mobile Business and Multilateral Security, Goethe University Frankfurt, Frankfurt am Main, Germany
david.harborth@m-chair.de

Abstract. Augmented reality (AR) greatly diffused into the public consciousness in the last years, especially due to the success of mobile applications like Pokémon Go. However, only few people experienced different forms of augmented reality like head-mounted displays (HMDs). Thus, people have only a limited actual experience with AR and form attitudes and perceptions towards this technology only partially based on actual use experiences, but mainly based on hearsay and narratives of others, like the media or friends. Thus, it is highly difficult for developers and product managers of AR solutions to address the needs of potential users. Therefore, we disentangle the perceptions of individuals with a focus on their concerns about AR. Perceived concerns are an important factor for the acceptance of new technologies. We address this research topic based on twelve intensive interviews with laymen as well as AR experts and analyze them with a qualitative research method.

Keywords: Augmented reality, qualitative user study, privacy concerns, security concerns, social consequences.

1 Introduction

Augmented reality (AR) gained much attention in the public in the last years, especially due to the success of mobile applications like Pokémon Go [1]. Since the origins of AR technologies in 1968 [2], there was a consistent technical development of AR hard- and software. However, the technical development of AR lacks accompanying research about user perceptions, attitudes and behaviors. Already in 2005, Swan and Gabbard [3] postulated the need to further develop AR systems from a technology-centric medium to a user-centric medium and demanded user-based experimentation. Comparable results are found by Harborth [4], who finds that the majority of research on AR deals with technical aspects and developments and only few user studies and evaluations are conducted. In addition, Dey et al. [5] report on ten years of user studies published in AR outlets. They find that most of these studies are formal user studies, with little field testing and almost no heuristic evaluations. The majority of existing

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

studies on augmented reality focuses on the behavior of users interacting with the technology, especially in laboratory environments.

Besides the prevailing gap in current research on AR, it is important to investigate this technology due to its potential impact on all aspects of individuals' daily lives. Tim Cook, CEO of Apple, states that *"[...] it will happen in a big way, and we will wonder when it does, how we ever lived without it. Like we wonder how we lived without our phone today"* [6]. This opinion is substantiated by an increasing activity in the mergers and acquisitions (M&A) market, where large technology companies like Apple or Snapchat are buying small specialized AR firms [7, 8]. Furthermore, AR is experiencing an increasing relevance in the B2B sector by providing substantial efficiency gains [9]. Thus, it is relevant to investigate the interaction with AR and the implications of using AR as early as possible. In summary, the current state of research and the importance of AR as an innovation make it necessary to conduct more user studies on AR in order to analyze users' needs and concerns. In contrast to quantitative research, a qualitative method makes it possible to investigate a wider horizon of possible perceptions and attitudes of individuals that previous research did not consider. Therefore, we investigate the perceived concerns of users with regard to AR technologies based on a qualitative method, called grounded theory method (GTM) [10]. The results can serve as a starting point to evaluate and foster market adoption as well as future developments of AR technologies.

The remainder of this paper is structured as follows. We review related work in Section 2. The methodology is described in Section 3. The results are presented in Section 4. Section 5 contains a discussion of the results. We conclude with Section 6.

2 Literature Review

Past and current research on AR is mainly technical. Comparable research investigating the potential and actual users of AR is relatively rare [3–5]. Most of the existing user studies follow a quantitative research method and investigate one specific case of an AR technology using known theoretical models and construct operationalizations like technology acceptance models or privacy concerns (e.g. [11–17]). Partially, studies complemented quantitative approaches with qualitative methods. For example, Grubert et al. [18] investigate user reviews for AR browser, focusing on negative comments and clustering these comments with regard to certain application features. Olsson and Salo [14] mix open questions with numerical evaluations in an online user survey to investigate general acceptance issues and user experiences with respect to mobile AR. Haugstvedt and Krogstie [13] investigate acceptance factors of a mobile AR application for cultural and historic information in cities. First, they employ a quantitative approach based on a technology acceptance model for hedonic information systems to test their adapted model. Second, they complement this analysis with qualitative interviews with users who actually interact with the application. Besides these mixed-method studies, we could only identify three studies that employ a pure qualitative research method. Ross and Harrison [19] assess technology acceptance factors of a specific AR application based on face-to-face focus groups. Olsson et al. [20] use contextual interviews (i.e. in the natural environment of the research object, in this case shopping centers) to investigate their expectations and requirements for mobile AR applications. The only study which applies a GTM is the one by Anuar [21] who investigates mobile AR apps for tourism and how such apps influence the experiences of users. Since our research does not focus on one specific application, it provides a higher level of analysis, indicating that the results hold for a variety of different AR technologies. In addition, perceived concerns of AR users are seldomly included in existing analyses. We identified only two articles dealing with one dimension of concerns for the case of Pokémon Go, namely privacy concerns. First, Harborth and Pape [11] specifically focus on the privacy concerns of actual players of the game. They find that although privacy concerns are relatively high, people are still playing the game. This indicates that benefits potentially outweigh the concerns. Therefore, as long as developers can address current concerns with regard to a missing added value of AR, it is possible that users will still adopt AR technologies in light of privacy concerns. The results by Rauschnabel et al. [22] indicate similar conclusions. They find that privacy risks have no significant effect on the attitude towards playing Pokémon Go. Besides these articles, other analyses of the different concerns could not be found.

3 Methodology

Our qualitative research approach relies partially on techniques of the GTM [10]. The GTM constitutes an inductive and comparative methodology for the development of theory grounded in systematic data collection and analysis. The theory involves an iterative process, moving back and forth between data gathering and data analysis; hence, it constitutes focused data gathering and makes the analysis more theoretical. At the center of grounded theory lies the constant comparative method, that is, new data is constantly compared with previous data to draw conclusions from similarities and differences within the data. We apply techniques from the constructivist grounded theory by Charmaz [23]. Constructivism assumes the existence of multiple social realities constructed by the research subject and the researcher and aims to interpret and understand the meaning of the constructed reality [8]. Since we think that interaction effects between researcher and interviewee are important to account for, we lean towards the constructivist GTM by Charmaz [23] in our research. GTM is a commonly used approach in the Information Systems (IS) domain for conducting qualitative research [24]. In a recent article, Wiesche et al. [24] provide a classification of 43 articles using GTM in the major IS outlets. The GTM is mainly supposed to generate a theory for a certain phenomenon under investigation. However, another purpose of GTM is to create a rich description of a phenomenon, in our case the possible concerns of individuals with respect to AR technologies. Since the GTM makes it possible to understand a certain problem from a variety of different perspectives, it is well suited for analyzing user perceptions about AR to generate specific recommendations for AR hardware and software developers to address these issues. We provide an overview of the single steps in the following sections.

3.1 Research Problem

As described earlier, the research objective of this paper is to analyze the individuals' concerns about AR technologies. For that purpose, we gathered data through intensive interviewing and created a semi-structured interview guide for conducting the interviews. Our interview guide is characterized by open questions and a high level of flexibility. This means, that we changed the interview guide according to new insights from already conducted interviews, e.g. by including interesting questions which came up in prior interviews. During the first interviews, our questions were rather generic. We asked interviewees about their personal experiences with AR, their beliefs about why AR is used, which application areas are favorable and how these use cases change in the future. Furthermore, we wanted to know the interviewees' assessments with regard to these future developments. When participants mentioned certain issues, barriers for adoption or concerns with regard to AR, we followed up on statements and investigated different dimensions of the statements in an explorative manner. The highly flexible approach with respect to the interview guide allowed for this procedure. Consequently, our questions were more guided during the last interviews when we focused on assessing whether our categories are saturated and whether they were covering the relevant sub-dimensions.

3.2 Data Collection Procedure

The requirements for including participants in the sample are the following. As the AR technology constitutes a new concept that is hard to compare with prior user experiences, the most important aspect for the initial sampling is that participants could answer certain questions correctly and, by that, showed either a certain degree of knowledge or actual experience with AR. Thus, we distributed questionnaires to potential interviewees to test this knowledge. We included interviewees in the sample when we could observe that this knowledge was sufficient. One of the interviewees is an expert in the field of AR with long-term experience in the field (interviewee 3).
After we conducted the first interview, the audio file was immediately transcribed and coded, and memos are written to reflect the discovered patterns and processes. Consequently, more data is gathered which focuses on the identified categories and their properties. This strategy is called theoretical sampling. Theoretical sampling aims to define the categories and their properties [23]. Our initial interviews include participants who exhibit diverse characteristics, including variation in age, gender, jobs, and distinct relations to AR, which we tested in the preliminary questionnaire. Our aim was not to create a sample that represents society, but to increase the potential variation within the data. We selected the participants according to two different strategies. The first strategy is to elaborate the new properties and dimensions of existing concepts based on participants who are similar to earlier interviewees (minimal comparison). The second strategy addresses the development of new categories and properties based on interviewees with different attributes (maximal comparison). The resulting categories, which are successively differentiated and enriched, thus control the further sampling process. The sampling process can stop when no new categories arise and the

properties of the identified categories are saturated. Saturated categories imply that gathering new similar and contrasting data does not lead to any new findings. Therefore, sample size is not the most important criterion when conducting a grounded theory research. For example, Bowen [25] states that a sample of 10 interviews done by a skilled interviewer could lead to more powerful insights than 50 interviews done by a novice. In this study, we conducted new interviews until no new categories or properties of categories emerged. This process led to a total of twelve analyzed interviews with a minimum length of 42 minutes up to 56 minutes (cf. Table 1).

Table 1. Demographics of the Interviewees

Interviewee	*Age*	*Gender*	*Education*	*Sector*
1	25-29	male	Bachelor	Consulting
2	30-34	male	Diploma	Consulting
3	45-49	male	Master	AR Industry
4	20-24	male	A Levels	Logistics
5	20-24	male	A Levels	Technical Services
6	25-29	male	Bachelor	Car Engineering
7	50-54	male	Diploma	Teaching
8	25-29	female	Bachelor	Information Mgt.
9	30-34	male	Master	Machine Construction
10	15-19	female	A Levels	Industrial Engineering
11	25-29	female	A Levels	Law
12	25-29	female	Bachelor	Managerial Economics

3.3 Initial and Focused Coding

In order to gain insights, to raise the analytical level of these insights and to construct theory from the gathered data, it is important to stop and analyze what has been found during the interviews. Analytic questions are asked to direct and further shape the following data gathering and to structure the data in a way that enables constructing theory. This step is called coding: Pieces of data are labeled to summarize and categorize the available data. Coding provides a tool to sort, structure and analyze a huge amount of data. For coding the data, we used the software package MAXQDA12. Grounded theory uses at least two phases for the data analysis, namely initial coding, in which the gathered data is analyzed line-by-line and segment-by-segment (initial codes fits the data closely); and focused coding, in which the most important and recurring initial codes are structured, integrated and arranged to gain analytical in-depth insights [23]. We used the most frequent initial codes that appear promising to reveal insights about concerns as focused codes. Where it was possible and provided analytic value, we aggregated different initial codes into a focused code that has a broader and more abstract view on the actions and their underlying patterns. After we defined a focused code, we compared it with previous and additional data to refine this code and to identify properties that can provide more insight.

3.4 Memo-Writing, Generation of Categories and Integration of Results

By applying the coding techniques, the data reveals certain structures and details. However, a link between the analysis of data in the coding phase and the creation of a theory or rich description that underlies the actions that have been identified remains absent. Memo-writing provides this intermediate tool and builds the bridge between the collection of data and a potential theory [23]. We used memos to summarize the statements of the interviewees, compare the statements made by participants to other statements in the same and prior interviews, structure patterns and processes of multiple interviews and to raise focused codes to conceptual categories.

The identified categories are the different dimensions of concerns. They represent the highest abstraction level of our analysis. They are defined by the mostly named codes which form the different sub-dimensions of this category (cf. Figure 1). As described earlier, these codes are closely linked to the data.

4 Results

An overview of the results is presented in Figure 1. The figure illustrates the abstraction process that emerged from applying the method described in the previous section. Categories represent the highest level of abstraction. We find that there are four dimensions of concerns of individuals with regard to AR technologies, which constitute the categories in our results. The dimensions are concerns about social consequences due to the use of AR, concerns about privacy, security and concerns about the missing added value AR might generate for the users. Figure 1 also shows the number of initial codes which form a sub-dimension (first number in parentheses). In addition, the second number shows how many different interviewees mentioned these codes. We will discuss each of the category in this chapter in more detail and provide exemplary statements from the interviews in Table 2.

4.1 Concerns about Social Consequences

This category is one of the most complex concepts due to the nature of AR, especially when considering HMDs. The first sub-dimension is the ***potential misuse of AR*** technologies. This focused code is closely linked to privacy concerns, but still different. The interviewees are not only concerned about the increasing possibility to gather data through the high pervasiveness of AR devices, but also about the potential consequences in their social life like a job loss. It is striking that the big benefits of AR, like constant analyses of the real environment, basically cause the concern about misuse scenarios. A strong ***involuntary exposure to ads ("Receiving Ads involuntarily")*** through AR devices is mentioned regularly amongst the interviewees. This concern is specific for HMDs considering the statement of interviewee 4. The other interviewees stated similar concerns about a constant flow of influencing advertisements. Considering the current state of internet marketing with an already high degree of ad exposure, it will be interesting to see how the relevant AR platform or device providers for HMDs deal with this situation.

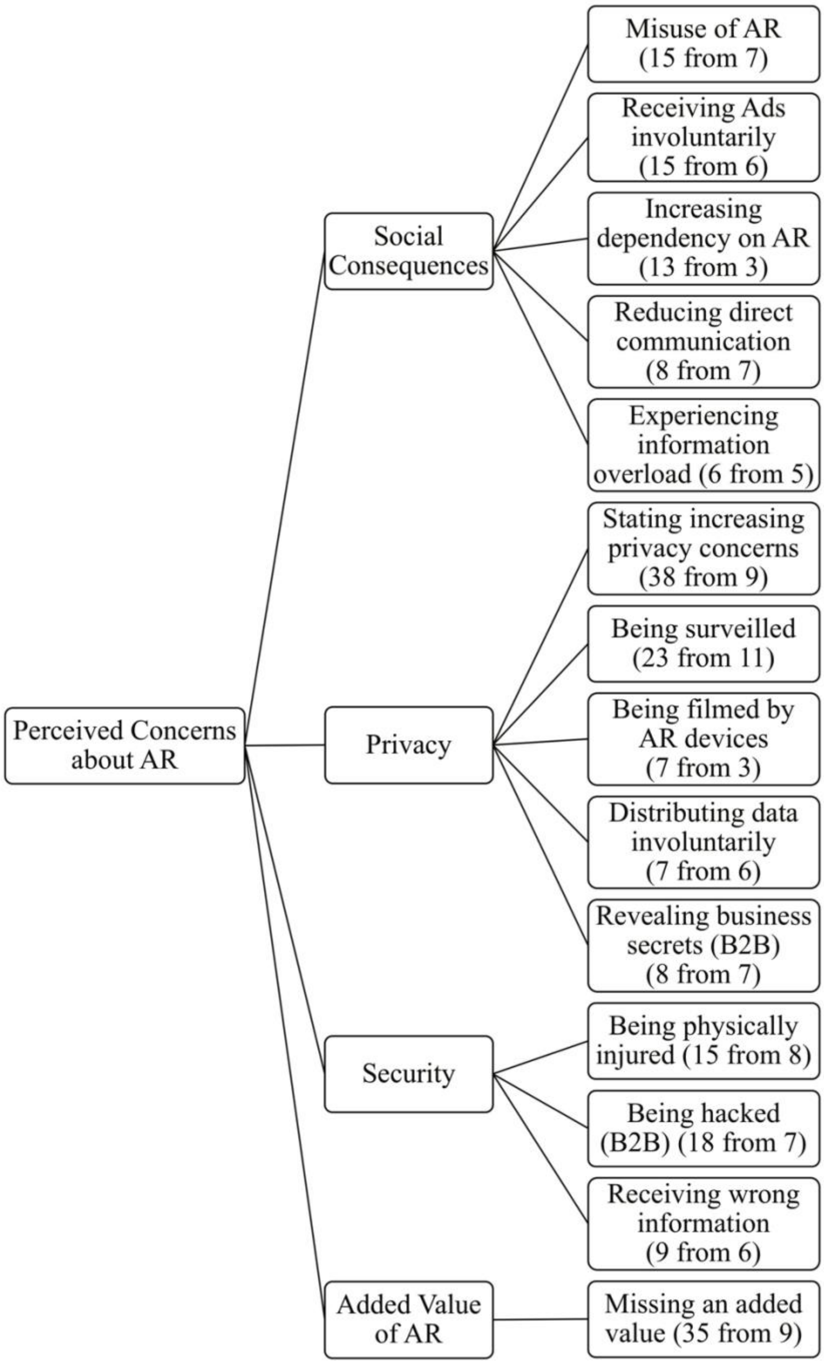


Figure 1. Results of the qualitative analysis process. The parentheses show the number of initial codes from different interviewees (e.g. the sub-dimension "being surveilled" is formed by 23 different initial codes, mentioned by eleven different interviewees)

It is promising to think about that form of advertisement from an economic point of view since users can be confronted with ads at the right place at the right time [26]. The third sub-dimension is about the ***increasing dependency on AR***. Our analysis reveals that the interviewees oftentimes refer to the smartphone when making certain judgements. Here, the exemplary statement also shows this comparison. Interviewee 7 sees a problem in a technology-dependent individual. For AR these concerns are apparently worse than for the smartphone. The next sub-dimension is related to the previous one. It emerges out of several interviewees saying that AR is most likely to ***reduce personal communication ("Reducing direct communication")*** amongst individuals even more than it happens already with the smartphone. This sub-dimension as well as the previous one are difficult to influence by AR developers or managers since they are problems inherent to several forms of AR devices like HMDs. The last sub-dimension is about concerns about ***experiencing an information overload*** when using AR. Information overload is a well-known problem in the digital sphere [27], encapsulated by the quote of Nobel laureate Herbert Simon when he said that "[...] a wealth of information creates of poverty of attention". Our analysis shows that individuals are rather not concerned about the sheer amount of data that is shown to them, but rather about the quality of information. Too much ”bad” information causes an information overflow since it is hard to differentiate which information is useful and which is not. This concern can be addressed by a well thought out information supply. Lastly, concerns about wearing HMDs in the social environment are not prevalent in this analysis. This is interesting since the media reported about incidents where people with the Google Glass were attacked by other people in their environment [28].

4.2 Concerns about Privacy

The literature provides several definitions of privacy and it can be said that Privacy is a concept in disarray. Nobody can articulate what it means [29]. In the context of this paper, a useful privacy definition is the one by Culnan, who states that privacy is the ability of an individual to control the access others have to personal information [30]. Our analysis of the privacy concerns matches the importance of control over personal information for individuals for several dimensions. The category privacy concerns is mainly about five sub-dimensions of concerns. First, interviewees perceive ***increasing concerns due to AR compared previous technologies*** like the smartphone. As interviewee 10 points out, this is mainly due to the perception that AR captures a larger amount and diversity of data. Closely linked to this, is the fear of ***being surveilled*** by government or other able parties. In contrast to this perspective, we find that interviewees also express concerns about the case where individuals ***being filmed*** (as bystanders) by an AR device worn by someone else. The next sub-dimension is about the ***involuntary data distribution ("Distributing data involuntarily")*** due to using AR. Besides following the prevalent law of the country, AR companies should provide transparency to show people what is being filmed, analyzed and stored to overcome these concerns. The last sub-dimension shows a different perspective by clearly focusing on business secrets. ***Being able to keep company information safe ("Revealing business secrets (B2B)")*** is apparently a prerequisite for a wide adoption of AR technologies in the business environment and need to be considered carefully when designing AR software and hardware.

4.3 Concerns about Security

Our analysis shows that security concerns can be divided into some kind of safety concerns, concerns about getting hacked (mainly mentioned in the B2B context) and concerns about receiving wrong information in the devices. Concerns about ***physical injuries ("Being physically injured")*** also came up in research articles and the press during first months after the appearance of Pokémon Go [31]. In our sample, interviewee 3 mentions potential long-term eye injuries due to the used technology. ***Concerns about hacks ("Being hacked (B2B)")*** are immanent when discussing security. The same holds for the possibility of ***receiving wrong information***. This is possible through malfunctions in the corresponding systems which can possibly lead to damaging exertions of influence.

4.4 Concerns about the Added Value of AR

The last category about the added value of using AR is not as fine-grained as the others. The only focused code we could identify is ***"missing an added value"***. However, it is a reoccurring theme across the interviewees that AR currently provides not enough added value because the technology is not mature enough. This evaluation of the interviewees is shared amongst several experts in the field of new technologies and AR. For example, Tim Cook, CEO of Apple, stated that "AR is going to take a while, because there are some really hard technology challenges there. But it will happen, it will happen in a big way, and we will wonder when it does, how we ever lived without it. Like we wonder how we lived without our phone today" [6]. Our results combined with such evaluations show that AR (especially in the form of HMDs) certainly needs more time and technical maturity to become a technology that is accepted in the end consumer market. In addition, the importance of showing a clear benefit of a certain AR technology to the customer has to be considered. Especially many mobile AR applications are rather designed to show the new possibilities of overlaying information over the real environment. However, only few of them provide a level of utility high enough in the users' daily lives so that they could represent more than just a shenanigan to once use for a few hours.

Table 2. Exemplary Quotes for the Sub-Dimensions from the Interviewees

Sub-Dimensions	Exemplary Statements
Misuse of AR (Section 4.1)	*And that [the data gathered through AR devices] could eventually be used against me sometime. If I say, I worked for my job. But then I was somewhere else and I used augmented reality to look something up, then it could actually be somehow used against me.* (interview 8) *Yes, so it could be [...] that my video recordings or my audio recordings or my behavior in principle by the use of augmented reality gets unlocked*

	for individuals or companies, where I usually do not really want, that they get this data. Be it any advertisers who use this data to send me personalized advertising or insurance, for example, who knows if I drive a little bit too fast in my car now and the augmented reality glasses will pass this data on to my insurer, who will then think about whether or not he will charge me the more expensive premium. (interview 1)
Receiving Ads involuntarily (Section 4.1)	*I would not want to be penetrated every day with any ads that I have to see. I mean, by that, I'm also influenced. That means, that the consumer is permanently influenced in the end.* (interview 4) *So that it now collects data about me, for example, what kind of clothes I normally wear, and when I walk past C&A, it then sends the data to me ah here you like to wear jeans or shirts and we just have shirts here on sale or something. So that would be where I would say, I do not want to see any ads unless I want them explicitly, but I cannot imagine that for my case.* (interview 2)
Increasing dependency on AR (Section 4.1)	*[...] the same positive and negative effects that had the smartphone, only reinforced. Even less self-responsibility, even more is transmitted, even more is automated, even less thinking [required by the individuals] in the negative sense.* (interview 7) *That somehow this [AR] manages this balancing act to have a noticeable advantage for me without that I feel so dependent on it.* (interview 8)
Reducing direct communication (Section 4.1)	*With them [smartphones] the social problems have increased. Social contacts are focusing a lot on the digital world and that social life, which normally exists, has already declined to a certain extent. And I also think Augmented Reality will reinforce that because it's not really necessary to meet friends, if you can see them in a live broadcast over the glasses, or yes, you can communicate directly via all possible communication media. And also has the feeling that you are sitting directly opposite a person.* (interview 12) *So everything that makes everyday life easier, you can clearly assume that that the society is getting lazy, more and more cumbersome. You can already see that. Um less communicates with each other, but more relies on devices. This is increasingly becoming a problem. So, you rely more and more on mechanics, more and more on technology and um, the trend is likely to continue.* (interview 4)
Experiencing information overload (Section 4.1)	*On the one hand, the data must come in smoothly, there must be no great delays, the data must be adapted, there must not be a flood of information, which is useless.* (interview 7) *Simple, that so to speak, that the reality gets into the background and just this information, this added information umm gets the upper hand and uh distracts you from possible events.* (interview 1)
Stating increasing privacy	*Now many see privacy as their private messages or at home private rooms, and through augmented reality you would be potentially recorded in these private rooms. And then perhaps, privacy is defined only for individual rooms, individual conversations or one has to say, that one is*

concerns (Section 4.2)	*private, if no one wears these glasses or so. Because I think that one already gives up parts of his privacy for it. Not that it might be unintentional or something, because we also use cell phones in our private rooms, but I think you could intervene in a different way [with AR devices]. Yes.* (interview 10) *And if you are in some kind of private sphere where you do not want visual or auditory information to leak out. I would probably still resign from it [using AR technologies] in this area until I am really sure that there prevails a sufficient privacy protection.* (interview 9)
Being surveilled (Section 4.2)	*[...] but most devices film the environment all the time, which is usually not recognizable for the surrounding people. And because of that, it would theoretically be possible to gather all the data and make a large-scale surveillance possible.* (interview 6) *Although we already have a lot of camera surveillance, you would also appear in the data of other people. So, by the fact that everyone could be monitored by anyone or just filmed.* (interview 10)
Being filmed by AR devices (Section 4.2)	*[...] if the technology would come to use in everyday life and [...] video material, for example, from different people walking around would be collected, then this would of course limit the privacy of the people according to current law* (interview 1) *For example, in my home I would not want a camera to permanently record my, yes my, apartment or my surroundings. My yes, my living situation or even if I am in the bathroom or something similar, that I am permanently filmed there.* (interview 12)
Distributing data involuntarily (Section 4.2)	*[...] what is recorded, is it recorded at all, how is it processed, where is it stored. I think that would be all the questions I would ask myself if I have such a device in my hand, because I'm very interested in the question, who can still see what I see?* (interview 5) *[...] I will probably never be able to control what it [the AR device] transmits, that is, which areas of my field of vision and whether it cuts out certain areas of my visual field so that people I see are not transmitted. I do not think we will have such options for an end user device. Basically, I would like to have a variety of granular options, for example, I want to say, I do not want data about other people to be transmitted or I do not want conversations with other people to be shared.* (interview 7)
Revealing business secrets (B2B) (Section 4.2)	*[...] um, because that's what the company lives on, that it's the only one who knows this information. That's why the problem is that you give the augmented reality application information so that it can process them furthermore, but this information must not be allowed to go outside. That's why I think companies need to establish a specific regulation where augmented reality can and cannot be used.* (interview 11)
Being physically injured (Section 4.3)	*Of course, there is much discussion with the glasses. What impact will this have on the eyes in the future? Umhm, above all new topics like light field technology, projections in the eyes and so on. Does it leave any damage in the long term?* (interview 3)

	There were also some cases in which you were so deepened in that [MAR applications] so you may also provoke accidents in the real world, because you do not see, for example, a lamp post, because somehow it is covered by other things. (interview 10)
Being hacked (B2B) (Section 4.3)	*Um, like all technical devices or even software, it is of course hackable, vulnerable.* (interview 2) *But in itself we would have the same concerns [as for smartphone and laptop use and storing data in data clouds] that our data could be passed on to third parties or these glasses could be hacked and this is currently also the case.* (interview 10)
Receiving wrong information (Section 4.3)	*[...] a set of algorithms behind it that can make mistakes, have wrong data, or possibly deliberately push in the wrong direction in the sense of propaganda. This is of course also possible if you are now really sitting at a data source and a large part of the people uses data similar to Google. That you can put your political mark on it, too.* (interview 7) *[...] gaps or IT-security deficits can be exploited very quickly and I think, examples would be the insertion of false information [...]* (interview 5)
Missing an added value (Section 4.4)	*So, if I refer it to myself now, [...] I think, as I said earlier, that it is not mature enough at the moment to really bring much benefit to me or change or facilitate my life drastically now, or so.* (interview 11) *[...] and above all, the question is what does it really bring me now? So, does it make sense now to buy a license for one hundred thousand euros to save a [previously used] device, but in the end it [AR glasses] shows me exactly the same information, but in a different form.* (interview 2)

5 Discussion

Our analysis provides a rich description of the different concerns expressed by the interviewees. We find that the main dimensions are concerns about social consequences, privacy, security and the missing added value through using AR. The results have practical implications for hardware and software developers as well as managers in AR companies with respect to specific design decisions that may weaken these concerns and foster future market adoption. As an example, our results indicate that the application of AR in the B2B context has to be accompanied with a well-thought policy on where the technology is allowed to be used and where it is forbidden. In the B2C context, AR products should be specified in a transparent way with respect to data gathering and processing. Furthermore, our research confirms the problem of the privacy of bystanders of AR users which is also found in the literature [32, 33]. Researchers, practitioners and governments have to find solutions to deal with this kind problem affecting the public sphere.

We contribute to the literature twofold. First, in contrast to the majority of earlier work on AR, we do not focus on technical issues of the technology, but solely on perceptions and concerns of potential users of it. Second, we show that individuals associate multi-faceted concerns with AR. These concerns are partially comparable to related findings about concerns with respect to smartphones. However, several sub-dimensions show AR-specific issues which have to be considered and investigated by researchers in order to overcome the concerns of potential users.

Since we applied the constructivist grounded theory approach by Charmaz [3], certain limitations arise. The method assumes a mutual construction of the reality by the researcher and the interviewee and it considers the researchers experience and view on reality. On the one hand, we choose the constructivist approach on purpose, because it supports our underpinnings and view on the world. On the other hand, we aim to minimize the personal influences on data collection to base our analysis in as much data as possible. However, our findings represent one possible version of the results which we analyzed based on our interactions with participants and the data. In addition, we did not set out to develop a model of how these concerns interrelate with each other and affect other variables like attitudes towards AR. Thus, this work only provides a rich description of the concerns related to AR as a first step to understand the perceptions of users in more detail.

Future work can build on our results to further investigate the most important concerns associated with AR technologies. A next logical step would be an extensive quantitative analysis of the influence of the different categories and dimensions of concerns on the intention to use AR, and if possible, on the actual use behavior. In addition, we could show that a qualitative approach is very useful for investigating innovative technologies like AR since most of the individuals are highly heterogenous with respect to their knowledge, perceptions and attitudes towards new technologies like AR. Thus, we recommend researchers to consider qualitative research methodologies for future user studies and evaluations to gain deeper insights into the users' perceptions and attitudes.

6 Conclusion

In this article, we identified four major categories of concerns about AR technologies. For every category, we disentangled the different sub-dimensions (if existent) and partly provided suggestions for how to deal with these concerns in order to increase the possibility of successful market entries of AR technologies in the future. Our analysis focused on AR technology in general and is therefore a first step towards a more granular understanding of concerns. Thus, the specific concerns can be more or less pronounced when looking at the specific type of AR technology. In addition, some of the sub-dimensions are not directly addressable by developers or managers (e.g. concerns about reducing direct communication or increasing dependency on AR) since they are, although very important, rather an object of a philosophical discourse that has to be conducted in the society itself.

7 Acknowledgements

This research was partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

References

1. Tassi, P.: “Pokémon GO” Finally Just Went Live On iOS And Android In The United States, http://www.forbes.com/sites/insertcoin/2016/07/06/pokemon-go-finally-just-went-live-on-ios-and-android-in-the-united-states/.
2. Sutherland, I.E.: A Head-mounted Three Dimensional Display. In: Fall Joint Computer Conference, Part I. pp. 757–764 (1968).
3. Swan II, J.E., Gabbard, J.L.: Survey of User-Based Experimentation in Augmented Reality. In: 1st Int. Conference on Virtual Reality. pp. 1–9 (2005).
4. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review. In: AMCIS. pp. 1–10. , Boston (2017).
5. Dey, A., Billinghurst, M., Lindeman, R.W., Swan, J.E.: A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. Front. Robot. AI. 5, 1–28 (2018).
6. Leswing, K.: Apple CEO Tim Cook thinks augmented reality will be as important as “eating three meals a day,” http://www.businessinsider.com/apple-ceo-tim-cook-explains-augmented-reality-2016-10?r=US&IR=T.
7. Loizos, C.: Snapchat has quietly acquired an Israeli startup for a reported $30 million to $40 million, https://techcrunch.com/2016/12/25/snapchat-has-quietly-acquired-an-israeli-startup-for-a-reported-30-million-to-40-million/.
8. Wolde, H. Ten: Apple buys German augmented-reality software maker Metaio, http://www.reuters.com/article/us-apple-metaio-idUSKBN0OE1RO20150529
9. Greenfield, D.: Augmented Reality for Industry, https://www.automationworld.com/article/industry-type/all/augmented-reality-industry.
10. 10. Glaser, B.G., Strauss, A.L.: The Discovery of Grounded Theory. Aldine Pub., Chicago (1967).
11. Harborth, D., Pape, S.: Privacy Concerns and Behavior of Pokémon Go Players in Germany. In: Hansen, M., Kosta, E., Nai-Fovino, I., and Fischer-Hübner, S. (eds.) The Smart Revolution. Privacy and Identity 2017. IFIP Advances in ICT, vol 526. pp. 314–329. Springer, Cham (2018).
12. Harborth, D., Pape, S.: Exploring the Hype: Investigating Technology Acceptance Factors of Pokémon Go. In: 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 155–168 (2017).
13. Haugstvedt, A.-C., Krogstie, J.: Mobile Augmented Reality for Cultural Heritage: A Technology Acceptance Study. In: 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 247–255 (2012).
14. Olsson, T., Salo, M.: Online User Survey on Current Mobile Augmented Reality Applications. In: 2011 ISMAR. pp. 75–84 (2011).
15. Rauschnabel, P.A., Brem, A., Ivens, B.S.: Who will buy smart glasses? Empirical results of two pre-market-entry studies on the role of personality in individual awareness and intended adoption of Google Glass wearables. Comput. Human Behav. 49, 635–647 (2015).
16. Rese, A., Schreiber, S., Baier, D.: Technology acceptance modeling of augmented reality at the point of sale: Can surveys be replaced by an analysis of online reviews? J. Retail. Consum. Serv. 21, 869–876 (2014).
17. Zahedi, F.M., Walia, N., Jain, H.: Augmented Virtual Doctor Office: Theory-based Design and Assessment. J. Manag. Inf. Syst. 33, 776–808 (2016).
18. Grubert, J., Langlotz, T., Grasset, R.: Augmented Reality Browser Survey. Tech. Rep. (2011).

19. Ross, H.F., Harrison, T.: Augmented Reality Apparel: an Appraisal of Consumer Knowledge, Attitude and Behavioral Intentions. In: HICCS Proceedings 2016. pp. 3919–3927 (2016).
20. Olsson, T., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. Pers Ubiquitous Comput. 17, 287–304 (2013).
21. Anuar, F.I.: Smartphone-Mediated Tourist Experiences: Understanding the Influence of Augmented Reality (AR) Applications in Tourism. Dr. Thesis, Texas A&M Univ. (2013).
22. Rauschnabel, P.A., Rossmann, A., tom Dieck, M.C.: An adoption framework for mobile augmented reality games: The case of Pokémon Go. Comput. Human Behav. 76, 276–286 (2017).
23. Charmaz, K.: Constructing Grounded Theory. Sage Publ., London (2014).
24. Wiesche, M., Jurisch, M.C., Yetton, P.W., Krcmar, H.: GROUNDED THEORY METHODOLOGY IN INFORMATION SYSTEMS RESEARCH. MIS Q. 41, 685–701 (2017).
25. Bowen, G.A.: Naturalistic inquiry and the saturation concept: a research note. Qual. Res. 8, 137–152 (2008).
26. Wassom, B.: Augmented Reality Law, Privacy, and Ethics: Law, Society, and Emerging AR Technologies. Syngress (2014).
27. Shapiro, C., Varian, H.R.: Information Rules: A Strategic Guide to the Network Economy. Harvard Bus. Sch. Press. (1999).
28. Weidner, J.B.: How & Why Google Glass Failed, http://www.investopedia.com/articles/investing/052115/how-why-google-glass-failed.asp.
29. Solove, D.J.: "I'Ve Got Nothing To Hide," and Other Misunderstandings of Privacy. San Diego Law Rev. 44, 745–773 (2007).
30. Culnan, M.J.: "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use. MIS Q. 341–364 (1993).
31. Howe, K.B., Suharlim, C., Ueda, P., Howe, D., Kawachi, I., Rimm, E.B.: Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. Bmj. 355, 1–4 (2016).
32. Denning, T., Dehlawi, Z., Kohno, T.: In situ with bystanders of augmented reality glasses. CHI '14. 2377–2386 (2014).
33. Rauschnabel, P.A., He, J., Ro, Y.K.: Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks. J. Bus. Res. 92, 374–384 (2018).

All websites have been accessed last on August 29th, 2018.

To (Psychologically) Own Data is to Protect Data: How Psychological Ownership Determines Protective Behavior in a Work and Private Context

Margareta Heidt[1], Christian Michael Olt[1], Peter Buxmann[1]

[1] Technische Universität Darmstadt, Information Systems, Darmstadt, Germany
{heidt,olt,buxmann}@is.tu-darmstadt.de

Abstract. The ever rising rates of data generation entail new opportunities for business and society but also an increasing risk of data breaches. Apart from technical measures, approaches like password authentication to ensure data protection revolve around the end-user as the human element in information security. Drawing on organizational research which argues that the sole feeling of ownership towards an intangible target like data can lead to heightened levels of the individual's responsibility, we investigate whether and to what extent this ownership feeling differs between personal files and data accessed in the work context. To this end, we draw on data derived through a two-phase questionnaire among a representative group of 209 employees. Consequently, we find evidence that psychological ownership shows stronger effects on protection motivation among participants in a private context. Furthermore, results indicate that employees partly relinquish their responsibility regarding security responses to protect data in their work context.

Keywords: password security, psychological ownership, employee, home user, protection motivation theory

1 Introduction

According to the latest estimations in 2012, 2.7 million terabytes existed in the digital universe with roughly 35 zettabytes of data generated annually by 2020 [1]. Data generation is further fueled through the acceleration of the Internet of Things and the growth of worldwide internet users to 4 billion in 2018 [2].

Unsurprisingly, the age of big data promises new opportunities for business and everyday life but entails new flip sides as evidenced by the ever increasing frequency and amount of damage of data breaches committed by cyber criminals. Verizon's annual report estimates that 81 percent of data breaches that occurred since 2014 were caused by stolen or weak passwords [3]. An estimation particularly striking given that the most prevalent approach to both access and protect private and business data remains through password authentication. Passwords can thus be considered a particular vulnerability as they are especially intertwined with the human element in information systems – the end-user. Since end-users have been continuously identified

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

as the "weakest link" within the security chain, behavioral information security research emerged as an important subfield of information systems (IS) [4, 5].

Research on human behavior in IS security has been drawing on psychology, criminology, or health science and various adapted frameworks and models have been applied within the end-user context, examining either employee or individual private user behavior [5-7]. These models show that factors such as the certainty of sanctions, the risk appraisal of a cyber threat or perceived behavioral control are strong indicators leading to the behavioral intention to perform certain protective actions [8, 9]. However, extant studies have only identified and analyzed the effectiveness of these factors on security in *either* a work environment *or* in the context of private use [10]. Thus, it remains unclear if certain factors affect the intention to behave in a more secure way in order to protect – one's own or the company's – data even though the context-sensitivity of findings has recently received increased attention among IS scholars [11].

In this regard, existing studies [e.g., 12, 13] have suggested that the sole feeling of possession or "being psychologically tied to an object" [14, p. 299] might lead to heightened levels of individual responsibility and engagement in IS security behavior. This feeling is referred to as "*psychological ownership*", a concept that describes the self-derived perception of ownership opposed to the actual legal ownership which is backed by the perception of others and the legal system. Psychological ownership (PO) is rooted within the innate human need to experience possession of either tangible or intangible targets [15] and the sense of regarding this target as extension of one's self [16]. In turn, human desire to experience control and accountability over the target differs according to the level of PO an individual experiences [14, 17].

However, IS studies thus far have either focused on feelings of ownership towards the targets 'internet' and 'one's computer' among home-users [12] *or* towards the target 'information' in a generic work-based scenario [13]. Whereas the first study argues for a direct influence of PO on intention to protect the target of ownership, the latter theorizes how PO affects the protection motivation, i.e., antecedents of intention to protect information. Due to the dearth of research on the influence of PO on security behavior, both aforementioned studies call for future research with Anderson and Agarwal specifically recommending additional studies that "explore the differences in behavior between employees and home users" [12, p. A15]. Nevertheless, other IS studies that integrated PO into the privacy calculus [18] or explored PO of IT [19] have only examined the role of PO in one single context and have not questioned yet how levels of PO might differ according to situational differences in contexts. But do individuals really experience the same degree of PO regarding, for example, their own electronic device or one provided through their company? Or do individuals experience higher levels of PO regarding their personal data as opposed to PO regarding the data they work with – and are supposed to protect through appropriate security measures – in their professional environment?

Against this backdrop, we seek to (1) extend prior IS research on individuals' protection motivation of data by highlighting the distinct role of PO both in a work and a private context. Furthermore, our study is the first to our knowledge that actually (2) compares protection motivation based on a repeated measures study design and one distinct sample in both contexts.

The remainder of this article is structured as follows: the theoretical background of both Protection Motivation Theory and psychological ownership is presented and serves as the foundation of our hypotheses which are integrated into a research model and tested in both a work and private setting. Subsequently, the results of our study are demonstrated and discussed before implications for theory and practice are derived.

2 Theoretical Background and Hypotheses Development

The following section provides an overview of the current state of behavioral IS security research in both work and private contexts along with the basics of the aforementioned concept of psychological ownership and how it has been accounted for thus far in IS security literature. Based on the theoretical background, hypotheses are developed and integrated into our research model which draws on Protection Motivation Theory (PMT).

2.1 Information Security Research

IS research has a long-standing tradition of analyzing security-related issues on an organizational and individual level [20, 21]. In an organizational context, researchers continue to advance technical approaches to prevent intrusion or to detect attacks [22, 23], however behavioral information security research has gained considerable momentum during the last two decades by focusing on human, and in particular, end-user behavior in work and private use contexts.

Within behavioral information security research, users in a work context can generally be divided into two subgroups: users that exhibit deviant behavior, i.e., compromising information security through espionage, theft, or sabotage, and those users who misbehave without the intent to cause damage [24]. By means of example, the latter group's misbehavior can manifest itself through defiance of security policy aspects such as using corporate devices to access non-work related websites or utilizing weak, repetitive and thus easy-to-compromise passwords for important work accounts [25]. In order to understand the driving factors of such "unintended" misbehavior or to identify aspects that encourage the use of safeguarding practices, IS researchers have heavily relied on behavioral theories that originate in behavioral psychology, organizational science, criminology, or health research [6-9, 12].

Protection Motivation Theory has been widely used to analyze "any threat for which there is an effective recommended response that can be carried out by the individual" [26, p. 409] and thus serves as a widespread theory in IS security research due to its applicability to security threats such as violating security compliances [27] or losing data due to irregular backups or weak passwords [28]. At the core of PMT, attitudes of individuals are assessed through two cognitive processes which lead to an increased intention to protect oneself against a potential threat: namely, *threat* and *coping appraisal*.

Threat appraisal comprises the perception and assessment of threat severity as well as the personal vulnerability to a threat. In our security context, perceived severity of a data breach and one's own vulnerability to fall prey to such an event will affect the protection motivation regarding data. As both perceived severity and vulnerability are positively correlated with the response behavior to protect one's data, which in our case will be through strong passwords, we hypothesize:

H1a *Perceived vulnerability will have a positive effect on an end-user's intention to protect (work and private) data.*
H1b *Perceived severity will have a positive effect on an end-user's intention to protect (work and private) data.*

Once the threat is assessed, e.g., the potential severity of data loss or theft and one's susceptibility or likeliness to experience such an incident, individuals will evaluate a potential behavioral response to the threat during the so-called coping appraisal.

Coping appraisal includes the concepts response efficacy and the associated response costs of the planned coping response necessary to protect oneself from the specific threat, as well as one's perceived self-efficacy in performing the response. If self-efficacy and response efficacy outweigh response costs, an individual yields a positive coping appraisal, i.e., individuals will install anti-virus software despite the associated costs in terms of purchase price or time to install because they feel capable of performing the installation and also deem the software to be effective in averting viruses and malware [12, 28]. More precisely, response efficacy in PMT refers to the belief that a certain response performed by the individual actually leads to a reduction or elimination of the considered threat.

Regarding IS security, end-users might wonder if strong passwords actually increase the security of their own or their company's data. If this specific response is considered effective in actually decreasing the threat (such as potential misuse of data caused by unauthorized access) an individual will be more inclined towards actually using strong passwords. However, this response also entails the cognitive effort of remembering several complex passwords. The concept of response costs thus assesses all efforts and expenditures associated with the coping behavior which will have a negative impact on the intention of actually performing the response in question. We thus hypothesize:

H1c *Response efficacy will have a positive effect on an end-user's intention to protect (work and private) data.*
H1d *Response costs will have a negative effect on an end-user's intention to protect (work and private) data.*

The core nomology of PMT additionally includes the concept of self-efficacy which has also been applied in various other theories to assess IS security behavior, often as part of the construct perceived behavioral control (PBC) [8, 29, 30]. On the one hand, self-efficacy relates to the confidence of individuals in their own skill, knowledge and ability to perform the response. On the other hand, controllability, as the second aspect of PBC, describes how much of the performance is actually up to the individual [30, 31]. One example would be that employees might be hindered to implement a security measure due to missing administrator rights on their work computers. Similar to other PMT-based studies which extended their research model with elements of the models originating from Theory of Planned Behavior, we also integrate the complete concept of PBC into our model as it could serve as a differentiator between the work and private context [12, 32].

In an IS security context, end-users who are confident in their ability to perform an appropriate security measure like backing up data at home or at their workplace, will be more inclined to progress with that chosen coping mechanism. However, controllability might differ across contexts, because employees might not express the same extent of assumed controllability regarding their actions if they cannot implement a security measure due to missing administrator rights – even if they had the skill and knowledge in doing so. As a result, they might shift the responsibility to their IT department or employer. Nevertheless, if employees just like private end-user ascribe responsibility to themselves, i.e., perceive higher degrees of controllability regarding the coping mechanism, they will be more proactive in taking appropriate security measures [32]. Hence, we expect that:

H1e *Self-efficacy will have a positive effect on an end-user's intention to protect (work and private) data.*
H1f *Controllability will have a positive effect on an end-user's intention to protect (work and private) data.*

2.2 Psychological Ownership

The following examples serve as an introduction to the general concept of PO: 1) Alice and Bob, both three-year-old toddlers, erupt in a fight over a doll in a physician's practice: both children claim the doll belongs to them and attempt to protect it from the other claimant by shouting "It is MINE!" – Although, technically, the doll is legally owned by the physician. 2) Alice's mother is a project manager. She lovingly calls one of her recent projects her 'baby' and takes many project-related tasks home to continue working after hours instead of delegating tasks because she feels a high sense of commitment and ownership towards this particular project. These scenarios depict how individuals behave when they feel that they possess an ownership stake in a physical or intangible object – a phenomenon called psychological ownership.

The term PO stems from psychology and describes the sense of ownership of a target like the aforementioned doll or project, but can also be felt towards a concept, another person, or an entire organization or community. The target is seen as an extension of the self [33], i.e., the owners regard the target as an expression of themselves or feel a strong sense of belongingness towards the target – as evidenced for example by football supporters who feel strong ownership towards their football club [14, 16]. Although related, PO is distinct from legal ownership which is recognized by society and protected by legislation – whereas PO is a "condition of which one is aware through intellectual perception […] coupled with an emotional or affective sensation" [34, p. 86]. The resulting effects of PO have been analyzed and categorized into positive outcomes – such as citizenship, personal sacrifice and assumption of risk, or experienced responsibility and stewardship – and negative effects like territoriality and other defiant behavior, or personal maladies like stress or frustration if the target is subject to any form of alteration [34, 36].

The roots of PO or the reason why this cognitive-affective state exists is best explained by an innate need of having a place or belongingness to the target [15], a sense of symbolic expression though the target or self-identity [16], and the desire to experience causal efficacy through control and accountability over the target [14, 17]. Due to the versatility of the PO concept, it has found extensive application especially in management and organizational research. More recently, studies in an IS context have started to introduce PO and demonstrated its impact on system usage and appreciation of IT [19, 35], willingness to disclose data [18], or intentions to perform security-related behavior [12, 13]. However, only the latter two studies examine the role of PO as antecedent of the threat and coping appraisal or its direct effect on intention in a behavioral security context [12,13]. In line with their reasoning and to further explicate how PO possibly affects the coping and threat appraisal, we present a third example: 3) Bob's father uses a CRM application on a daily basis at work which is accessed through password authentication. He is fully aware of the criticality of customer and business data stored in the application and has an intimate knowledge of many entries since he found and recorded a lot of the information himself. The data is not simply his company's data but also his own in his perception and a loss thereof would hurt him personally. Thus, threats to the target can be regarded as threats to oneself because the target represents an extension of one's self-concept or identity. In our context, higher levels of PO will lead to heightened perceptions of severity and vulnerability when faced with the prospect of losing one's data. This will more likely occur in the context of private data as opposed to data in a work context. Hence, we assume that risk appraisal will be influenced through psychological ownership as follows:

H2a *PO of (work and private) data will increase perceptions of threat vulnerability. This effect will be more pronounced in a private context.*
H2b *PO of (work and private) data will increase perceptions of threat severity. This effect will be more pronounced in a private context.*

Intimate knowledge or a deep understanding and familiarity of an object will lead to higher degrees of association with the object [37]. This is evidenced by individuals' statements of preferring own targets to comparable others, simply because one knows them better, e.g., the favorite spot in the canteen. Acquiring knowledge about a target is also linked to investment of the self into the target which represents the third route to PO. Investing time, effort, or energy into the creation or development of a target, e.g., in a mentor-mentee relationship or into do-it-yourself-projects, facilitates feelings of PO by seeing one's own reflection in the target [34]. In organizational studies, employees who feel PO toward their company are shown to express higher levels of organizational commitment, organizational-based self-esteem, and job performance [36, 38, 39]. Subsequently, Pierce and colleagues argue that pronounced feelings of PO will influence the degree of its effects – both positive and negative [38]. In line with Menard and colleagues, we also expect that PO will exert influence on the coping appraisal considering the use of diverse and strong passwords in both the private and work context [13], and thus hypothesize:

H2c *PO of (work and private) data will increase perceptions of response efficacy. This effect will be more pronounced in a private context.*
H2d *PO of (work and private) data will decrease perceptions of response costs. This effect will be more pronounced in a private context.*

Apart from intimate knowledge of the target, and investment of the self, Pierce and colleagues also argue that perceived control is closely tied to feelings of PO [14]. Numerous studies prove that control is a core feature of ownership as objects that are habitually used or can even be manipulated by an individual become more assimilated into the user's self-concept [17]. According to Avey and colleagues, individuals will be "feeling more efficacious about working with the target, feeling more accountable for what happens with respect to the target" [39, p. 24] when they feel psychologically tied to the target. In our context, PO regarding their data will thus facilitate feelings of responsibility and as a result lead to heightened levels of willingness and confidence in their ability to carry out a protective response against the IS security threat [40].

H2e *PO of (work and private) data will increase perceptions of self-efficacy. This effect will be more pronounced in a private context.*
H2f *PO of (work and private) data will increase perceptions of controllability. This effect will be more pronounced in a private context.*

3 Research Model and Methodology

Our research model draws primarily on the approach of Menard and colleagues [12] which examines how psychological ownership affects the protection motivation based on PMT. We further extend the model with the additional construct of controllability in order to include and examine another important but yet often overlooked aspect of perceived behavioral control. In both contexts, we examine how the behavioral intention to use strong passwords in order to protect data is influenced by both the classic determinants of PMT and how these are in turn influenced by PO in our two contexts.

3.1 Data Collection Procedure

In order to investigate our research questions, we conducted an online survey among employees in Germany who use electronic devices to access software applications or websites and are interacting with company data in their professional environment on an everyday basis. Since comparing work and private contexts involves repeated measures for our research model, certain biases have to be considered and countered. In order to avoid that participants remember their answers from the first context (i.e., work) and aspire to repeat the same answer in the second context (i.e., private) to ensure self-consistency [42], we decided not to survey both contexts (work vs. private) in one single questionnaire. Furthermore, we chose to survey the same panel of respondents (cohorts) for both contexts. This enables us to investigate differences in an individual's perception handling work-related data as well as private data.

As threat scenario we chose the misuse of data caused by insecure passwords. Consequently, the coping strategy depicted the usage of strong passwords which are distinct between different user accounts. Both questionnaires were distributed online in August 2018 in two waves with the help of a market research institute. Seven days upon completion, the same cohort was invited to participate in a second survey assessing their password behavior within their private context. This timespan was chosen in order to avoid manipulating risk appraisal and coping appraisal between both conditions through unforeseen incidents or factors (work vs. private context) and is comparable to other IS studies using a repeated measures design [41, 42]. Using an even longer time-span entails the risk of external influences (security incidents and other message cues) to bias the respondents' perceptions compared to the first wave. Both surveys commenced with a welcome page which ensured the participants' anonymity and that there are no "wrong" answers in order to counteract common method and social desirability biases [13,43].

Only those participants who completed both survey questionnaires were included in the data analyses (N = 217). Since eight participants failed our attention check during the second wave, their answers were deemed unreliable resulting in a final sample size of 209. The effective response rate after the second wave – and the elimination of unreliable responses and attention checks – amounted to 70.37 percent, an acceptable rate for questionnaires considering security-related behavior [e.g., 28, 44].

The sample was evenly distributed in terms of gender (51.2 % female; 48.8 % male) and age (mean = 44.9; min = 19; max = 65) through quotas mirroring the percentage of the overall population in Germany and thus providing an adequate snapshot of reality of German employees. We report a more detailed sample statistic in the online appendix (Table A1).

3.2 Operationalization of Research Variables and Instruments

All measurements to operationalize our research variables are based on previously validated operationalization and have been adapted to the context of our study as we report in the online appendix (Table A2). The items for all threat related PMT constructs (vulnerability [VULN], severity [SEV]) were adopted from Johnston and Warkentin [45]. Items for response efficacy [RE] have been extracted from Witte [46] whereas response costs [RC] as well as self-efficacy [SE] were adapted from Milne at al. [41]. Controllability [CON] was measured using the scale from Kraft and colleagues [47]. Our dependent variable behavioral intention [INT] has been operationalized using items from Herath and Rao [25] whereas psychological ownership [PO] has been adopted from van Dyne and Pierce [38].

4 Data Analysis and Results

Our data set contains 209 responses for each context based on the same respondent cohort. Therefore, we distinguish between two contexts: the work versus the private context. In the following, the hypothesized relationships between variables are analyzed relying on the PLS algorithm as implemented in SmartPLS in order to simultaneously validate the measurement model and the conceptual path model [48].

Measurement Model Testing. We begin by assessing convergent validity of all our variables for each condition (work and private). Internal consistency can be assumed for constructs if Cronbach's alpha (Cr α) as well as composite reliability (CR) are at least 0.7 [49]. To establish convergent validity, the average variance extracted (AVE) should exceed 0.5 [50]. In addition, item loadings are assessed against a threshold of 0.65 or higher [51]. We find minimum loadings of 0.707 / 0.840 in the work and private context respectively. Therefore, we conclude that convergent validity is ensured.

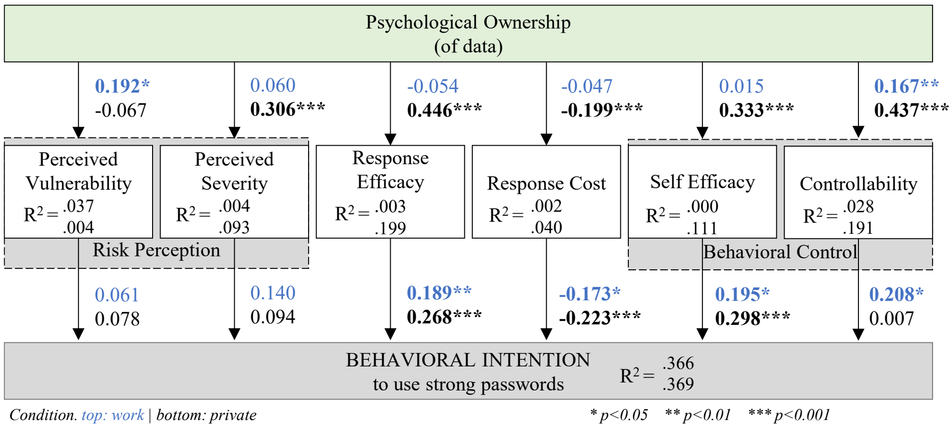
Table 1. Measurement Model Validation

	Work Context										
	Cr α	CR	AVE	CON	INT	PO	RE	RC	SE	SEV	VULN
CON	.915	.937	.790	**.889**							
INT	.948	.967	.906	.383	**.952**						
PO	.906	.941	.843	.167	.095	**.918**					
RC	.940	.957	.848	.252	.403	-.054	**.896**				
RE	.879	.924	.803	-.248	-.399	-.047	-.223	**.921**			
SE	.854	.912	.775	.236	.469	.015	.387	-.662	**.880**		
SEV	.927	.948	.821	.273	.355	.060	.361	-.126	.315	**.906**	
VULN	.794	.859	.606	.004	.001	.192	-.061	.237	-.116	.102	**.778**
	Private Context										
	Cr α	CR	AVE	CON	INT	PO	RE	RC	SE	SEV	VULN
CON	.937	.954	.840	**.916**							
INT	.930	.955	.876	.366	**.936**						
PO	.888	.931	.818	.437	.297	**.904**					
RC	.949	.963	.867	.582	.364	.446	**.898**				
RE	.881	.926	.807	-.354	-.446	-.199	-.070	**.931**			
SE	.881	.927	.810	.399	.514	.333	.189	-.746	**.900**		
SEV	.929	.950	.825	.252	.260	.306	.364	-.080	.137	**.909**	
VULN	.898	.928	.762	-.235	-.102	-.067	-.171	.356	-.209	.095	**.873**

For acceptable discriminant validity, we rely on the criteria suggested by Fornell and Larcker [52]. Accordingly, the square-root of AVE (**bold** numbers in Table 1) needs to be greater than the correlations to all other constructs. Since this holds true for all constructs within both conditions, we assume our measurement model to be accurate as further evidenced by cross loadings reported in the online appendix (Table A3).

Structural Model Testing. Continuing with the validated measurement model, we assess the overall model fit of our conceptual models. The standardized root mean square residual (SRMR) is 0.066 resp. 0.046 (work resp. private) which is well below the cutoff-point of 0.08 recommended by Hu and Bentler, indicating a good model fit [53]. The amount of variance explained within our dependent variables (R^2) are presented in Figure 1. We use a bootstrapping procedure with 5,000 subsamples to test for statistical significance of path coefficient estimates which results are also reported.

Figure 1. Results of the PLS Model Estimation



Work context. In the work context, all hypotheses based on PMT (H1c, d, e, f) except for risk perception, i.e., perceived vulnerability and severity, are supported. Furthermore, psychological ownership of data shows only significant influences on the variables perceived vulnerability (H2a) and controllability (H2f).

Private context. The results show that risk perception has no significant influence on behavioral intention to use strong passwords in the private context. Nevertheless, response efficacy, response costs and self-efficacy significantly influence behavioral intention to use strong passwords and thus support H1c, d, e. The expected effect of controllability on behavioral intention is not supported. However, PO has a strong influence on the majority of PMT related constructs (H2b, c, d, e, f supported) whereas the effect on perceived vulnerability could not be supported in the private context (H2a rejected).

Multi Group Analysis. As an extended analysis of the differences between the two contexts, we conducted a multi group analysis. Due to space limitations we report hypotheses which ultimately show significant differences in their path-coefficients only in our online appendix (Table A4). Hereby, the context shows a mediating effect on H1f and H2a as the effects are stronger in the work context compared to the private context. The context furthermore mediates all other relations from psychological ownership. Hence, the effect of psychological ownership is stronger in the private context compared to the work context for H2b, c, d, e, f.

5 Discussion and Contributions

With this study, we contribute to a better understanding of PMT according to the situational context and via an extension through PO. By using a repeated measures design, we are able to demonstrate varying mechanisms leading to the intention to protect either private or work data through strong passwords. Specifically, our results demonstrate that risk appraisal through perceived severity and vulnerability does not significantly affect the intention to use a security measures such as strong passwords which is in line with some recent findings of other researchers [8, 13].

Furthermore, we find significant differences regarding the effect of controllability across contexts: whereas a significant effect of controllability on the intention to use strong passwords indicates that employees feel accountable for their choice, this effect could not be shown among private end-users. This might indicate that they do not even perceive an opportunity to shift control, and thus accountability, to some third party such as the employer. Therefore, we find evidence that individuals in the private context are aware of their sole accountability when responding to security threats. Otherwise, we found the influence of coping appraisal to be generally stronger in a private context.

Similarly, but opposed to the study of Menard and colleagues, we could demonstrate lesser and mostly insignificant effects of PO on PMT antecedents in a work context [13]. PO effects are mostly only significant in a private context apart from the hypothesized influence on perceived vulnerability – which, in turn, is only evident in a work context – and with controllability which is significant in both contexts. Additionally, a post-hoc performed paired t-test ($t(208) = -20.36$; $p < 0.001$) of PO according to the condition work ($M = 3.07$; $SD = 2.77$) or private context ($M = 5.89$; $SD = 1.34$) showed significant differences. Accordingly, we can subsume that PO is more pronounced considering the protection of private data and, as individuals tend to evaluate a target more favorably when they own it, feelings of accountability, responsibility, and investment of the self in the target are stronger [34, 38, 39]. This leads to several potential implications for both theory and practice.

Theoretical Contributions. From a research point of view, our approach is the first to our knowledge that is based on a repeated measures design which enables the comparison of PMT's explanatory power in a work and private context based on the same safeguarding behavior, i.e., the use of strong passwords. Our study contributes to an improved understanding of the relationships within the theory and shows varying support of the general concepts of risk and coping appraisal. Risk perception in isolation does not promote safeguarding measures in any context, whereas the inclusion of controllability could contribute to more thorough understanding of employee intention regarding the use of strong passwords. Additionally, our findings contribute to the still scarce literature on psychological ownership in IS security. IS research and studies on information security in particular, have incorporated PO very rarely and diversely in terms of context and the mode of influence which calls for replication studies as called for by Menard et al. or Anderson and Agarwal [12, 13]. In this regard, we could demonstrate that PO significantly influences several PMT antecedents only in a private context and barely affects the protection motivation among employees.

Practical Implications. As such, our study informs IS scholars but also practitioners about how a sense of ownership can regulate protection motivation and thus lead to the actual use of safeguarding mechanisms like strong passwords. Practitioners in particular should stimulate feelings of PO regarding company data in order to increase protection motivation. PO can, for example, be increased and stimulated by tapping into its antecedents, e.g., through more intimate knowledge of the target, in our case data. Also if employees invest more time and effort into understanding how data can be protected, employees will develop a feeling of freedom of choice and more accountability which in turn increases PO and thus exert a positive effect on safeguarding mechanisms [19].

6 Conclusion, Limitations, and Future Research

Despite taking all necessary measures to ensure qualitative results, our study is not without limitations. In this regard, a typical limitation of behavioral IS theories is the measurement of intention rather than actual behavior. Although intention is widely regarded to be a very robust predictor of actual behavior [29, 54], future research could build onto our findings with an experimental design that observes the influence of PO on actual behavior. Similarly, previous research has identified several other influencing factors like culture or personal characteristics which had to be omitted due to duration constrains but could enhance our understanding about the modes of action of PO in an organizational and individual security context. Especially, since culture has been shown to have an effect of the level of PO expressed, our results could be culturally constrained to Western, more individualistic, cultures [13]. From a methodological point of view, the rather short timeframe of our two surveys might add to the general finding that humans strive for a consistent manner of self-representation which might result in memory effects or so-called experimenter demand effects [42, 43]. However, an extension of the time frame might be affected by unidentifiable external influences due to unforeseen incidents or other biases that arise during the survey period.

An avenue for future research could be the analysis of PO antecedents through an action research design measuring whether increased feelings of PO also lead to improved actual security behavior in both a work and a private context. Our study thus serves as an important stepping stone which first compared the behavior of individuals in these contexts in a repeated measures design revealing varying degrees of effect sizes in well-established PMT and newly hypothesized PO relationships. Furthermore, future research could develop a new operationalization of PO for the IS context, as current measures are often based on physical, tangible targets. A different approach could be the use of an Implicit Association Test which can detect underlying attitudes of users or consumers particularly when subjects are unaware or unwilling to identify sources of influence – like PO in our context [55]. This could prove to be particularly interesting since an additional estimated PLS model including a path from PO to intention showed that no direct influence on intention was found in both contexts as opposed to the previous studies. This could be related to the differences in targets as one's own device might elicit more pronounced feelings of PO compared to intangible data or the operationalization of PO through scenario manipulation and represents another possible avenue for future research [11].

7 Acknowledgments

This research project was funded by “Hessisches Ministerium des Innern und für Sport”. We express our gratitude to everybody involved in enabling us with the conduct of this study – in particular to Henrik Schlünkes and Zhengchao Wu.

References

1. IBM: The Flood of Big Data - Driving Marketing Effectiveness by Managing, http://www.ibmbigdatahub.com/infographic/flood-big-data (Accessed: 08.08.2018)
2. Statista: Global Digital Population as of July 2018, https://www.statista.com/statistics/617136/ (Accessed: 08.08.2018)
3. Verizon: Verizon Data Breach Investigations Report, https://www.ictsecuritymagazine.com/wp-content/uploads/2017-Data-Breach-Investigations-Report.pdf (Accessed: 08.08.2018)
4. Schneier, B.: Secrets and Lies: Digital Security in a Networked World. John Wiley and Sons., New York, NY (2000)
5. Crossler, R.E., Johnston, A.C., Lowry, P.B., Hu, Q., Warkentin, M., Baskerville, R.: Future Directions for Behavioral Information Security Research. Comp. & Sec. 32, 90-101 (2013)
6. Lebek, B., Uffen, J., Neumann, M., Hohler, B., Breitner, M.: Information Security Awareness and Behavior: A Theory-based Literature Review. Management Research Review 37, 1049-1092 (2014)
7. Mayer, P., Kunz, A., Volkamer, M.: Reliable Behavioural Factors in the Information Security Context. 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy (2017)
8. Boss, S.R., Galletta, D.F., Lowry, P.B., Moody, G.M., Polak, P.: What do Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Behaviors in Users. MIS Quarterly 39, 837-864 (2015)
9. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Information Security Policy Compliance: an Empirical Study of Rationality-based Beliefs and Information Security Awareness. MIS Quarterly 34, 523-548 (2010)
10. Mou, J., Cohen, J., Kim, J.: A Meta-Analytic Structural Equation Modeling Test of Protection Motivation Theory in Information Security Literature. 38th International Conference on Inf. Systems (ICIS), Seoul, South Korea (2017)
11. Davison, R.M., Martinsons, M.G.: Context is King! Considering Particularism in Research Design and Reporting. Journ. of Inf. Technology 31, 241-249 (2016)
12. Anderson, C.L., Agarwal, R.: Practicing Safe Computing: A Multimedia Empirical Examination of Home Computer User Security Behavioral Intentions. MIS Quarterly 34, 613-643 (2010)
13. Menard, P., Warkentin, M., Lowry, P.B.: The Impact of Collectivism and Psychological Ownership on Protection Motivation: A Cross-Cultural Examination. Comp. & Sec. 75, 147-166 (2018)
14. Pierce, J.L., Kostova, T., Dirks, K.T.: Toward a Theory of Psychological Ownership in Organizations. Academy of Management Review 26, 298-310 (2001)

15. Duncan, N.G.: Home Ownership and Social Theory. In: Duncan, J.S. (ed.) Housing and Identity: Cross-Cultural Perspectives, pp. 98-134. Croom Helm, London (1981)
16. Dittmar, H.: The Social Psychology of Material Possessions: To Have Is to Be. Hemel Hampstead, Harvester Wheatsheaf and St Martin's Press, New York (1992)
17. Furby, L.: Understanding the Psychology of Possession and Ownership. Social Behavior and Personality 6, 49-65 (1978)
18. Cichy, P., Salge, T.O., Kohli, R.: Extending the Privacy Calculus: The Role of Psychological Ownership. 35th International Conference on Information Systems (ICIS), Auckland (2014)
19. Klesel, M., Ndicu, M., Niehaves, B.: Exploring Psychological Ownership of IT: An Empirical Study. 24th European Conference on Inf. Systems (ECIS , Istanbul, Turkey (2016)
20. Straub, D., Welke, R.: Coping with Systems Risk: Security Planning Models for Management Decision Making. MIS Quarterly 22, 441-469 (1998)
21. Liang, H., Xue, Y.: Understanding Security Behaviors in Personal Computer Usage: A Threat Avoidance Perspective. Journ. of the Association for Inf. Systems 11, 394-413 (2010)
22. Hansen, J.V., P.B., L., R., M., McDonald, D.M.: Genetic Programming for Prevention of Cyberterrorism through Dynamic and Evolving Intrusion Detection. Decision Support Systems 43, 1362-1374 (2007)
23. Cavusoglu, H., Raghunathan, S., Cavusoglu, H.: Configuration of and Interaction between Information Security Technologies: The Case of Firewalls and Intrusion Detection Systems. Inf. Systems Research 20, 198–217 (2009)
24. Willison, R., Warkentin, M.: Beyond Deterrence: An Expanded View of Employee Computer Abuse. MIS Quarterly 37, 1-20 (2013)
25. Herath, T., Rao, H.R.: Protection Motivation and Deterrence: A Framework for Security Policy Compliance in Organisations. Europ. Journ. of Inf. Systems 18, 106-125 (2009)
26. Floyd, D.L., Prentice-Dunn, S., Rogers, R.W.: A Meta-Analysis of Research on Protection Motivation Theory. Journ. of Applied Social Psychology 30, 407-429 (2000)
27. Pahnila, S., Siponen, M., Mahmood, A.: Employees' Behavior towards IS Security Policy Compliance. 40th Annual Hawaii International Conference on System Sciences (HICSS), Hawaii, US (2007)
28. Crossler, R.E.: Protection Motivation Theory: Understanding Determinants to Backing Up Personal Data. 43rd Annual Hawaii International Conference on System Sciences (HICSS), Hawaii, US (2010)
29. Ajzen, I.: The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes 50, 179–211 (1991)
30. Ajzen, I.: Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. Journ. of Applied Social Psychology 32, 665-683 (2002)
31. Bandura, A.: Self-Efficacy: Toward a Unifying Theory of Behavioral Change. Psychological Review 84, 191-215 (1997)
32. Workman, M., Bommer, W.H., Straub, D.: Security Lapses and the Omission of Information Security Measures: A Threat Control Model and Empirical Test. Computers in Human Behavior 24, 2799-2816 (2008)
33. Webb, T.L., Sheeran, P.: Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence. Psychological Bulletin 132, 249-268 (2006)
34. Pierce, J.L., Kostova, T., Dirks, K.T.: The State of Psychological Ownership: Integrating and Extending a Century of Research. Review of General Psychology 7, 84-107 (2003)
35. Lee, Y., Chen, A.N.K.: Usability Design and Psychological Ownership of a Virtual World. Journ. of Management Inf. Systems 28, 269-307 (2011)
36. Vandewalle, D., Van Dyne, L., Kostova, T.: Psychological Ownership: An Empirical Examination of its Consequences. Group & Organization Management 20, 210-226 (1995)

37. Rudmin, F.W., Berry, J.W.: Semantics of Ownership: A Free-recall Study of Property. Psychological Record 37, 257–268 (1987)
38. Van Dyne, L., Pierce, J.L.: Psychological Ownership and Feelings of Possession: Three Field Studies Predicting Employee Attitudes and Organizational Citizenship Behavior. Journ. of Organizational Behavior, 439-459 (2004)
39. Avey, J.B., Avolio, B.J., Crossley, C.D., Luthans, F.: Psychological Ownership: Theoretical Extensions, Measurement and Relation to Work Outcomes. Journ. of Organizational Behavior 30, 173-191 (2009)
40. Dipboye, R.L.: A Critical Review of Korman's Self-Consistency Theory of Work Motivation and Occupational Choice. Org. Beh. and Human Perf. 18, 108-126 (1977)
41. Milne, S., Orbell, S., Sheeran, P.: Combining Motivational and Volitional Interventions to Promote Exercise Participation: Protection Motivation Theory and Implementation Intentions. British Journ. of Health Psychology 7, 163-184 (2002)
42. Kehr, F., Kowatsch, T.: Quantitative Longitudinal Research: A Review of IS Literature, and a Set of Methodological Guidelines. 23rd European Conference on Inf. Systems (ECIS), Münster, Germany (2015)
43. Podsakoff, P.M., MacKenzie, S.B., Leong-Yeon, L., Podsakoff, N.P.: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. Journ. of Applied Psychology 88, 879-903 (2003)
44. Sonnenschein, R., Loske, A., Buxmann, P.: Gender Differences in Mobile Users' IT Security Appraisals and Protective Actions: Findings from a Mixed-Method Study. 37th International Conference on Inf. Systems (ICIS), Dublin, Ireland (2016)
45. Johnston, A.C., Warkentin, M.: Fear Appeals and Information Security Behaviors: An Empirical Study. MIS Quarterly 34, 549-566 (2010)
46. Witte, K.: Predicting Risk Behaviors: Development and Validation of a Diagnostic Scale. Journ. of Health Communication 1, 317-342 (1996)
47. Kraft, P., Rise, J., Sutton, S., Røysamb, E.: Perceived Difficulty in the Theory of Planned Behaviour: Perceived Behavioural Control or Affective Attitude? British Journ. of Social Psychology 44, 479-496 (2005)
48. Bagozzi, R.P., Yi, Y.: On the Use of Structural Equation Models in Experimental Designs. Journ. of Marketing Research 26, 271-284 (1989)
49. Bagozzi, R.P., Yi, Y.: Specification, Evaluation, and Interpretation of Structural Equation Models. Journ. of the Academy of Marketing Science 40, 8-34 (2012)
50. Hair, J.F., Ringle, C.M., Sarstedt, M.: Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. Long Range Planning 46, 1-12 (2013)
51. Falk, R.F., Miller, N.B.: A Primer for Soft Modeling. University of Akron Press, Akron, Ohio (1992)
52. Fornell, C., Larcker, D.F.: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journ. of Marketing Research 18, 39-50 (1981)
53. Hu, L.-T., Bentler, P.M.: Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. Psychological Methods 3, 424-453 (1998)
54. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading, MA (1975)
55. Brunel, F.F., Tietje, B.C., Greenwald, A.G.: Is the Implicit Association Test a Valid and Valuable Measure of Implicit Consumer Social Cognition? Journ. of Consumer Psychology 14, 385-404 (2004)

Understanding Data Protection Regulations from a Data Management Perspective: A Capability-Based Approach to EU-GDPR

Clément Labadie[1], Christine Legner[1]

[1] Faculty of Business and Economics (HEC), University of Lausanne, Switzerland
{clement.labadie, christine.legner}@unil.ch

Abstract. The European General Data Protection Regulation (EU-GDPR) has entered into force in May 2018. Its emphasis on individual control and organizational accountability constitutes a new paradigm that requires changes in the way organizations manage personal data. However, organizations face difficulties when implementing EU-GDPR due to a lack of common ground between legal and data management domains. Anchored in the resource-based view theory (RBV), this paper argues that the regulation requires companies to build a dedicated data management capability. It presents a capability model that was developed in an iterative design science process, integrating both interpretation of legal texts and practical insights from focus groups with more than 30 experts and from 3 EU-GDPR projects. The paper advances the regulatory compliance management literature by translating legal data protection concepts for the IS community. It also contributes to practice by enabling organization to set-up systematic approaches towards EU-GDPR compliance.

Keywords: EU-GDPR, Data Protection, Regulations, Compliance, Capabilities.

1 Introduction

In 2017, The Economist published an article entitled "The world's most valuable resource is no longer oil, but data" [1], mirroring the transformation of our modern economies, in which massive data collection and analysis have become a key competitive advantage. This transformation had led the European Union (EU) to start a major reform of its data protection framework, which resulted in the adoption of the General Data Protection Regulation (EU-GDPR) in 2016, and its enforcement in May 2018. The EU-GDPR constitutes a paradigm shift in data protection, towards greater choice and sovereignty for individuals, and more accountability for organizations [2]. For organizations, it comes with the burden of proof related to whether, how and how well they protect personal data and increased fines for noncompliance. This requires them to fundamentally rethink the way they store and process personal data on an

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

enterprise-wide level. Despite the past deadline, most companies have not yet reached full EU-GDPR compliance. A study conducted in April 2018 among more than 1000 European and US companies reported that 40% of respondent organizations would not comply on May 25th, 2018. And even if companies have started to address GDPR, only 23% of US-based companies and 31% of EU-based companies stated that they were confident with their ability to comply [3].

The difficulties in implementing EU-GDPR highlight the general lack of common ground between legal and IS in both research and practice. From the research side, legal aspects of information privacy were not among the "topic areas closer to the interests of most IS researchers" [4], and the few IS studies on EU-GDPR have a very restricted scope. Similarly, in most companies, data protection topics have traditionally been addressed by legal departments by adapting contracts and general conditions, but without directly influencing data management practices. However, the new regulation does not allow for such a restricted approach, and companies see data processing related issues as the most challenging topics in EU-GDPR. In fact, preparing for data breach notification, operationalizing data portability, operationalizing the right to be forgotten and conducting data inventory/mapping were cited as "most difficult GDPR obligations to comply with" [3]. Furthermore, our interactions with practitioners indicate that the regulation is very generic, and that there is a need to translate it into data management concepts and practices. This "translation" would help analyze compliance requirements and options, before deciding on concrete (technical) implementations.

Anchored in the resource-based view theory (RBV), this paper argues for utilizing capabilities as an interface between abstract compliance requirements and their concretization. It aims at addressing the following research question: what data management capabilities need to be built in order to address EU-GDPR's requirements? Following a design science research approach, we propose a capability model for EU-GDPR that integrates both interpretation of legal texts and practical insights from focus groups with experts from 22 companies as well as 3 EU-GDPR related projects. The resulting capability model comprises organizational and system capabilities from a data management perspective. In contrast with the few existing research papers on EU-GDPR that treat selected aspects of the regulation, such as data breach notification or data portability, our study thereby provides an integrated perspective on enterprise-wide data management practices. The resulting capability model may also act as a classification framework for those studies that investigate specific aspects of the regulation.

The remainder of this paper is structured as follows: we first introduce the EU-GDPR as well as an overview of current research on the topic and on regulatory compliance in general. After outlining the research methodology and process, we motivate the capability perspective and present the capability model. We conclude by summarizing our contribution and discussing future research.

2 Background and related research

2.1 The European General Data Protection Regulation (EU-GDPR)

In January 2012, the European Commission published a proposal for an overhaul of data protection law within the European Union, which would become EU-GDPR[2]. It thereby addressed the need to remedy the fragmented implementations of the preceding Data Protection Directive (95/56/EC), as well as to account for the significant changes introduced by the internet and digital services [5], [6]. As a result, EU-GDPR directly applies in every EU member state. Moreover, any organization that processes personal data of EU-citizen must comply with it, regardless of the geographical location of their operations. If it fails to do so, fines with significantly heightened amounts will apply (i.e., up to 20 million euros or 4% of an organization's global revenues, whereas previous regulations averaged at ca. 500 000 euros). EU-GDPR reinforces existing concepts, and introduces new ones. Most notably, existing transparency mandates have been strengthened – organizations must now inform individuals about data processing in clear language and separately from general conditions, and are also required to present more granular consent options [5]. One of the major additions is the concept of accountability, which implies that organizations must be able to demonstrate compliance with the regulation. They must also appoint data protection officers (DPOs) and announce data breaches to both authorities and individuals (data breach notification). Privacy-by-design principles (i.e., implementing privacy from the ground up in systems and offerings) also appear in the regulation, along with new individual rights, such as data portability as well as a right to oppose automated decision making [5]. All of these evolutions constitute a paradigm shift in data protection, towards greater choice and sovereignty for individuals, and more accountability for organizations [2], [6], [7].

2.2 EU-GDPR and Data Protection in IS Literature

Although EU-GDPR was finalized in 2016 and presents a major paradigm shift in data protection, it has attracted relatively little attention in IS literature so far. A query with the keyword "GDPR" returns 27 results on the AIS Electronic Library, as the time of writing (September 2018). The majority of these papers simply mention EU-GDPR, but only seven studies treat it as key topic. From Table 1, we see that existing EU-GDPR studies fall in the domains of information privacy practices (5 studies) and information privacy technologies and tools (2 studies), in [4]'s taxonomy of topic areas. However, with the exception of [14], all studies exclusively focus on one of EU-GDPR's requirements. There are two shortcomings in this approach: First, none of them is aimed at analyzing the entire regulation and its implication from an enterprise-wide perspective. Second, these papers do take the compliance requirements for granted and directly look into specific practices. Hence, we are still lacking a broader understanding

[2] Regulation (EU) 2016/679. Recitals (R.) and articles (art.) mentioned throughout the text refer to EU-GDPR unless otherwise specified.

of the challenges faced by companies in implementing EU-GDPR. [14] addresses this topic by proposing a Digital-Privacy Transformation "Gap-Map" that measures the organization's propensity for change. However, it exclusively takes a change management perspective, without investigating the compliance requirements and their implications on enterprise-wide data management practices.

Table 1. Summary of EU-GDPR-Related Studies in IS Literature

	Study type	EU-GDPR aspects	Topic area based on [4]	Level of analysis	Research focus
[9]	Empirical	Data breach notification	Information privacy practices	Orga-nisation	Applying data breach notification to past infringements
[10]	Conceptual	Data breach notification	Inf. privacy practices	Orga-nisation	Information security / incident management
[11]	Conceptual	Data portability	Information privacy impact	Market	Impacts of data portability right on competition dynamics
[12]	Conceptual	Privacy-by-design	Technologies and tools	Indivi–dual	Privacy label for GDPR
[13]	Conceptual	Tran-sparency	Technologies and tools	Orga–nisation	Guidelines for com-pliant privacy notices
[14]	Conceptual	Entire regulation	Impact / Inf. privacy practices*	Orga–nisation	Transformation framework for digital privacy
[8]	Empirical	Account-ability	Inf. privacy practices	Market	Review of third-party data processors
** scope beyond [4], covering organizational and individual readiness and transformation*					

2.3 Regulatory Compliance Management (RCM)

So far, the academic discussion on EU-GDPR has not linked up to the regulatory compliance management (RCM) research domain, although the latter could inform how to analyze regulations and their influence on business practice. RCM is defined as "ensuring that enterprises are structured and behave in accordance with the regulations that apply, i.e., with the guidelines specified in the regulations" [15]. RCM introduces useful background definitions to delimit relevant legal concepts. In his overview paper, [15] distinguishes between regulations (i.e., binding document), regulatory guidelines and compliance requirements, as provided in the legal text. Following interpretation, this ultimately results in concretized compliance requirements as implementation.

Two review papers from 2009 have analyzed the coverage of RCM in IS research. [16] conducted a literature analysis through the lens of enterprise architecture, and isolated 26 relevant papers. They found that while some aspects of RCM have been prominently studied (e.g. organizational and behavioral impacts of regulations, compliance supporting IT solutions), others had been neglected. Specifically, they found no contributions on the operationalization of compliance objectives. [17]'s literature analysis on RCM, revolves around the approaches (i.e., explanatory or solution) and context (i.e., region, type and domain) of the considered contributions. From the 45 papers, the majority focused on North America, whereas only 3 of them focused on European issues. Related to data protection, they identified 2 papers that study Fair Information Practices, and only one on the European Data Protection Directive (95/46 EC), even though it had been enforced for 15 years. Furthermore, all identified contributions offered either preventive or detective solutions, but no corrective solutions. The authors hypothesize that corrective solutions are an outcome of legal analysis, which is why they were not addressed by the IS community.

Hence, there is a lack of RCM-related contributions that address data protection regulations, focus on regions other that North America [17] and provide guidance to concretize strategic compliance objectives [16]. This last call is echoed by our literature review on EU-GDPR – although contributions exist around the topic, they all focus on specific aspects of the regulation, and lack a single integrating framework.

3 Research method

Given our stated goal to support companies in achieving EU-GDPR compliance, we adopt design science research (DSR) to develop a capability model, as an artefact "to solve identified organizational problems" [18]. Table 2 depicts the research steps, following the iterative process suggested by [19] and outlines the close interactions between academics and practitioners, comprising 5 focus group meetings with 33 data management experts from 22 companies and insights from 3 EU-GDPR projects.

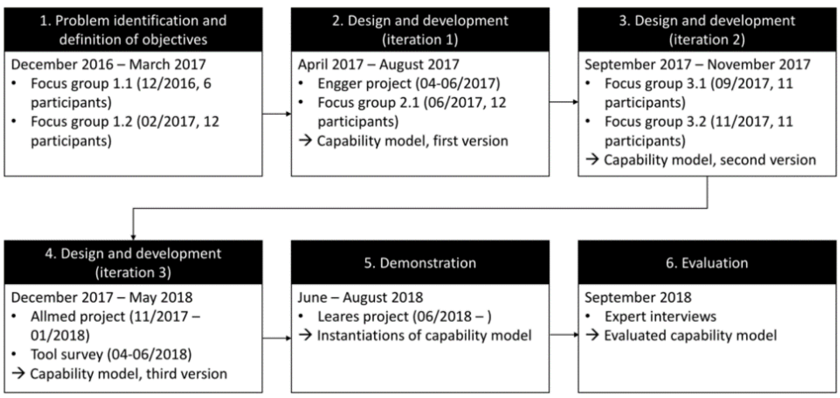


Figure 1. Research process (based on DSRM)

The **first phase** was meant to understand the problem at hand and specify the objectives of the solution to be developed. It was conducted between November 2016 and March 2017 based on an initial review of the regulation, with the objective of isolating requirements relevant for data management. We started by extracting and analyzing EU-GDPR's compliance requirements according to foundational data protection principles in legal literature (i.e. personal data, informational self-determination, accountability and transparency). To that end, we selected reference text books that provide a comprehensive analysis of data protection foundations and applications – they integrate legal texts and their related preparatory works, as well as insights from case law and legal doctrine [20–22]. Early results of this analysis were discussed with practitioners through focus groups 1.1 and 1.2, allowing them to reflect on the regulation's impacts on their organizations and implementation challenges. These discussions led to an in-depth understanding of the issues in implementing EU-GDPR enterprise-wide and the subsequent decision to design a capability model.

The next phases (2, 3 and 4) were **iterative design cycles**, involving insights from field projects and internal research activities to design the capability model, as well as focus groups for collecting feedback. Internal research activities included a continuous analysis of EU-GDPR-specific legal literature [2], [5–7], [20], [22], [23], guidelines from official authorities [24–26] as well as interpretations from the private sector, including consortia (e.g. [27–30]) and industry stakeholders (e.g. [31]).

Phase 2, the first design iteration phase, comprised a project at Engger[3], a global engineering company, and resulted in the initial version of the capability model. It had just started a multi-project around EU-GDPR-compliant personal data aiming at harmonizing business partner data management in a highly distributed landscape, i.e., with around 500 systems in different countries and subsidiaries. This project helped understanding issues and define capabilities related to collection and distribution of personal data and consent. It ultimately led to the first version of the capability model that was presented to and discussed with data management experts in focus group 2.1.

During **phase 3**, the discussions in the two focus group meetings 3.1 and 3.2 revolved around the scope of the model. Feedback from focus group 3.1 indicated that security is usually a distinct function, and supported the need for a data management-centric perspective. From an academic perspective, information security is a well-research field and the existing concepts may be translated to EU-GDPR, whereas there is little coverage of data management practices in regulatory compliance with data protection regulations. It was decided to set aside all security-related considerations from the capability model and focus exclusively on data management capabilities.

Phase 4 comprised a project around consent management at Allmed, a global pharmaceutical company. Its technical team had designed an MVP solution, which we analyzed based on the second version of the capability model. Insights from the project resulted in the capability model's third and final version. Afterwards, we analyzed software tools from major vendors claiming to support EU-GDPR compliance – to that end, we designed a taxonomy of desired functionalities based on the capability model (following the methodology proposed by [32]) and used it to classify 23 tools from major vendors. This analysis allowed us to validate the system capabilities.

Phases 5 and 6 included a demonstration with the EU-GDPR activities at Leares, a small consulting firm. The capability model proofed to be applicable and useful for assessing the current capabilities, identifying the required capabilities and prioritizing compliance activities. Additional expert interviews confirmed utility of the artefact.

[3] All company names have been anonymized.

4 Data Management Capabilities for EU-GDPR

4.1 Problem Formulation and Definition of Objectives

Discussions held in focus groups 1.1 and 1.2 revealed two main challenges with regards to GDPR compliance. First, participants recognized a lack of understanding of the regulation itself, while anticipating significant changes to the current way of storing and processing personal data on an enterprise-wide level. Second, they cited a lack of common ground with legal departments. In their organizations, discussions around data protection and privacy regulations are often cut short due to a lack of common approaches and vocabularies, which blocks the identification of feasible and compliant solutions and hinders progress. This led to the research objective of defining a capability model for EU-GDPR that assists data management professionals to understand and implement the regulation, as well as collaborate with legal colleagues.

4.2 Capabilities as a Link Towards Concretized Compliance Requirements

As theoretical foundation, we chose to rely on the RBV, as regulatory compliance is a component of firm performance, and contributes to an organization's control objectives (as defined by [33]). Building on [34]'s definition of an IT capability, we define data management capabilities for regulatory compliance as a firm's ability to acquire, deploy, and leverage its data resources in combination with other resources and capabilities in order to achieve an organization's compliance objectives.

Table 2. Positioning capabilities within RCM concepts

RCM concept	**Definition (based on [15])**	**Illustration in EU-GDPR**
Regulatory guideline	Stipulates a set of obligation to comply to.	Art. 6 – "Lawfulness of processing": enumerates conditions in which data processing is legal.
Compliance requirement (CR)	Pieces of text extracted from the regulatory guideline specifying an expected behavior / a specific condition to fulfill.	Extraction of requirements bearing data management relevance. E.g. art. 6 § 1 a and art. 7 § 1 require that data be processed according to individuals expressed consent.
Capability	*Result of the interpretation of CRs in terms of capabilities that are to be implemented or improved.*	*Manage consent and sub-capabilities: implement consent items, collect consent instances, distribute consent, enforce consent-based processing.*
Concretized compliance requirement (CCR)	Implementation of a CR in an enterprise model, fulfilling its legal specification.	A concrete measure implemented in a specific organization to operationalize CRs. E.g. "In company X, consent data should be first recorded in system 1 and pushed to other systems every 12 hours".

The capability model complements RCM concepts [15] and acts as an abstraction layer between the normative aspects of the regulation, i.e. the regulatory guidelines and compliance requirements (CR), and the concretized compliance requirements (CCR), i.e. the concrete implementation of a CR. Introducing capabilities allows describing what organizations should do, as opposed to how they should do it, i.e. how the specific implementation should be carried out. Table 2 depicts this articulation.

4.3 Capability Model: Structure and Overview

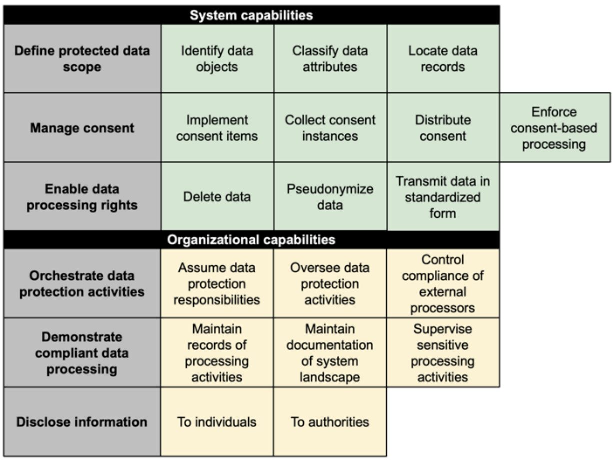


Figure 2. Capability model for data management in EU-GDPR

Art. 24 § 1 states the overall responsibility of organizations with regards to the regulation as the implementation of "appropriate technical and organizational measures to ensure and be able to demonstrate that processing is performed in accordance with this Regulation". We thus derived our two main capability groups, i.e., system and organizational capabilities (see Figure 2), reflecting their predominant aspect[4]. Correspondingly, **system capabilities** are mainly enabled by data-processing systems, while **organizational capabilities** rely on data protection processes and responsibilities. Capabilities were derived from EU-GDPR's underlying principles, as described by legal literature, and reflect the "pillars" of the regulation. Sub-capabilities are the result of the analysis and express compliance requirements. In the following sections, we present each of the suggested capabilities, along with its justification, the empirical evidence and the sub-capabilities.

Define Protected Data Scope. This capability is based on art. 1 § 1 and 4 § 1 and denotes the ability to clearly identify, classify and locate personal data. Personal data is defined as "data enabling direct or indirect identification of a single physical person,

[4] In the RBV, capabilities "involve complex patterns of coordination between people and between people and other resources" [35]. Authors relying on the RBV in the IS literature usually demarcate technological and organizational aspects that underpin IS capabilities [36, 37].

data that is specific to a single physical person without enabling identification, data that can be linked to a physical person, data regarding which anonymization techniques cannot completely mitigate the risk of re-identification” [21].

Focus groups 1.1 and 1.2 indicated that companies generally had no overview on the personal data collected and used during processes, especially in terms of storage location. A participant of focus group 3.2 asked: “How do you identify personal data in a heterogeneous IT-System landscape?” Follow-up questions revolved around means to identify personal data. The project at Engger provided significant insight regarding this capability group. One of its main objectives was making sure that personal data was consistently kept up-to-date within all systems, which proved difficult due to multiple overlapping systems managed in independent subsidiaries. Overall, companies faced two main challenges: determining what kind of personal data they were processing, and where such data was stored. The resulting capability may be best summarized by [20], stating that “organizations must have perfect knowledge of personal data”. Practitioner reports also fall in line with this statement – [29] recommend two actions that mirror these issues (e.g. data discovery and system mapping) and suggest that personal data should not only be identified, but also classified. This is required as EU-GDPR prescribes higher protection levels for data that is considered sensitive (R. 51). The resulting sub-capabilities are:

- **Identify data objects:** identify data domains and related data objects that fall within EU-GDPR’s scope of applicability.
- **Classify data attributes:** assign levels of sensitivity to data attributes contained within personal data objects.
- **Locate data records:** identify all storage instances of personal data objects and have the ability to access and retrieve them.

Manage Consent. This capability comprises the prerequisites for collecting consent and ensuring consent-based processing of information. The principle of consent [5], [20, pp. 12, 94], [22, p. 93] is arguably one of the pivotal concepts of EU-GDPR and an expression of the right to informational self-determination. It can be defined as ability for each individual to determine whether and to what ends information about themselves can be processed [6]. The related concepts of conditionality, granularity and specificity are the most challenging for data management [24]. Conditionality (art. 7 § 4) means that consent for processing activities cannot be bundled in general conditions, and that a difference should be made between necessary and optional processing activities for a given purpose. Granularity (R. 43) implies that each processing activity and related consent item must be presented separately. Specificity prescribes a 1:1 relationship between processing types and consent items (i.e. yes/no question that relates to a personal data processing activity). This is a departure from practices before GDPR, when consent was mostly obtained through the bulk acceptance of general conditions.

Consent management found a significant echo in our focus groups. During focus group 3.1, none of the participants reported solutions either in final stages nor operational. During focus group 3.2, more questions were asked regarding consent management than all other capabilities combined. The Allmed project goal was making consent information accessible and readable by all systems, which mirror capabilities "distribute consent" and "enforce consent-based processing". However, difficulties arouse in two areas. First, the system would need to be connected to every system storing and processing personal data – identification of such systems proved difficult and the existing system landscape documentation was deemed insufficient (see the capability "define protected data scope"). Second, the team struggled to identify consent items, as they were usually contained in unstructured form (e.g. within general conditions, contracts, webpages). A specific sub-capability was added to reflect this issue, and is a prerequisite to all other consent-related capabilities. The resulting sub-capabilities are:

- **Implement consent items:** define and implement consent items that mirror data processing activities performed throughout business processes.
- **Record consent instances:** collect and record consent expressed by individuals.
- **Distribute consent:** ensure consent items updates in all affected processing systems.
- **Enforce consent-based processing:** ensure that data processing activities are performed in accordance with consent expressed by individuals.

Enable Data Processing Rights. This capability denotes the ability to process data according to EU-GDPR's data rights and principles. It was derived from the principle of accountability (art. 24 § 1), but covers only the technical aspects to reach compliance, document them, and provide proof of compliance [5], [20, p. 12], [22, pp. 31, 38].
Art. 17 provisions a "right of erasure", according to which individuals can request that organizations delete their personal data (provided that they have no other obligation to keep said data). From a technical perspective, enterprise systems usually prevent users from deleting data and practitioners expressed a difficulty in that regard. When asked about it, none of the participants of focus group 3.1 reported that they had operational deletion processes or mechanisms. Focus group 3.2 also expressed a lack of well-established solutions at this level, and our tool study identified only 2 solutions supporting this capability. Art. 25 mandates privacy by design / by default approaches, including the principle of minimization [22, p. 90], i.e. processing as little personal data as possible. One way of operationalizing it is pseudonymization, which is a rare occurrence of EU-GDPR mentioning a specific technological approach (R. 28-29). This can be seen as an alternative to deletion, as pseudonymized data exits EU-GDPR's scope of applicability, and was thus added as second order capability. Art. 20 introduces a "right do data portability" – organizations are required to transmit personal data records "in a structured, commonly used and machine-readable format" to individuals, and, in some cases, directly to other organizations. During focus group 3.1, only a quarter of respondents declared that the provision of data in standardized formats was mature, and none of them reported working communication channels. We have identified only two solutions in our tool study, both of which are marketed as "Customer Identity and Access Management Systems" (CIAM). The resulting sub-capabilities are:

- **Delete data:** permanently remove data records from their systems.
- **Pseudonymize data:** use pseudonymization techniques in order to adhere to the principle of minimization.

- **Transmit data in standardized form:** transmit personal data to external parties using standard formats and set up communication channels with other organizations.

Orchestrate Data Protection Activities. This capability denotes the organizational ability to coordinate and execute data protection activities, involving different roles and responsibilities. It was derived from the organizational component of the principle of accountability [5], [20, p. 12], [22, pp. 31, 53, 71]. As stated, focus group feedback indicated that data managers often are at a loss as of who to consult when faced with data protection inquiries. This became particularly clear during the Allmed project – when the team needed to obtain information regarding data protection matters, they did not have a clearly designated contact person. On several occasions, responsibilities (e.g., for defining consent items) were not clearly defined. Art. 37-39 requires that organizations of a certain size appoint a "Data Protection Officer" (DPO). The DPOs should monitor compliance by acquiring an overview of processing activities, serve as advisory contact person [25], oversee record keeping and cooperation with authorities. We designed related capabilities for data protection oversight.

EU-GDPR also makes a distinction between data controllers and processors, and art. 28 orders the former to control compliance of the latter. This distinction is relevant to organizations when they outsource data processing to third party companies – the use of cloud services also falls into this situation, as merely storing data is considered processing. This became apparent during the Allmed project (cloud CRM) and especially in the case of Leares, which exclusively relies on cloud services (e.g. CRM, content management, websites) for the storage and processing of data. A corresponding capability was therefore added. The resulting sub-capabilities are:

- **Assume data protection responsibilities:** responsibilities for data protection-related tasks in all business functions that routinely process personal data.
- **Oversee data protection activities:** a leading role should oversee, organize, control and coordinate data protection activities.
- **Control compliance of external processors:** monitor that data processing conducted by third parties for EU-GDPR compliance.

Demonstrate Compliant Data Processing. This capability comprises the ability to record and evaluate sensitive processing activities, as well as to document system landscapes. It was derived from the documentation component of the principle of accountability [5], [22, p. 44]. Art. 30 orders organizations to "maintain a record of processing activities under its responsibility" and details the contents of such documentation. It was identified as a significant difficulty by [30], and all participants of focus group 3.2 acknowledged that documentation represented a significant effort. Maintaining system landscape documentation was identified as another sub-capability, as focus groups indicate that most organizations have difficulties locating data – this was the very motivation for the Engger project, and one significant roadblock for Allmed's solution implementation. Art. 35-36 further require organizations to conduct and document in-depth data protection impact assessments (DPIA) when performing sensitive processing activities. The resulting sub-capabilities are:

- **Maintain records of processing activities:** inventory and document personal data-related activities performed throughout business processes.
- **Maintain documentation of system landscape:** inventory and document systems that store and process personal data on a regular basis.
- **Supervise sensitive processing activities:** identify and evaluate sensitive data processing activities.

Disclose Information. This capability involves the ability to disclose information to individuals (R. 58) and authorities (art. 31). It was derived from the principle of trans–parency, which requests data protection measures to be clearly exposed [5], [20, p. 17].

Transparency requirements apply in two cases [26]. First, at the point of data collection, organizations must present related information separately, in a manner (e.g., language, illustrations) that can be easily comprehended. This would include, for instance, a clear description of each consent item. Transparency also refers to communications with individuals after data is collected, when organizations are faced with right-related requests (e.g., access, rectification, deletion). Art. 31 specifies that organizations "shall cooperate, on request, with the supervisory authority in the performance of its tasks". This implies that organizations set up a contact person for authorities (usually the DPO), and the ability to present relevant information / documentation as proof of compliance.

These capabilities may be seen as the operationalization of the principle of accountability, which is materialized by documentation. Since such documentation should contain all relevant information regarding an organization's data protection practices, these capabilities are about presenting that information to the interested parties (i.e. individuals and authorities). The resulting sub-capabilities are:

- **Disclose information to individuals:** provide individuals with complete and understandable information regarding the processing of their personal data and respond to their data protection-related requests.
- **Disclose information to authorities:** collaborate with designated data protection authorities and communicate relevant information upon request.

4.4 Demonstration

The main purpose of the capability model is to guide organizations in implementing EU-GDPR's requirements into their existing practices. To demonstrate its applicability and usefulness in EU-GDPR initiatives, we present how it was applied to assess the situation of Leares, a small-sized consulting firm that had started to draft a GDPR "action plan". A lengthy to-do list compiled the most visible and pressing compliance issues (e.g. adapting web forms, newsletters and contracts) in order achieve what was considered a "minimum" level of compliance. There were significant shortcomings with this approach. First, there was no indication of why certain actions were necessary, or what compliance issue they were meant to fix. Second, actions were presented as isolated, one-time efforts – there was no indication as to what extent GDPR compliance was actually achieved, or how it would be maintained in the future. Third, and most notably, these action items focused mostly on technical issues, with no documentation mechanisms or compliance processes put in place.

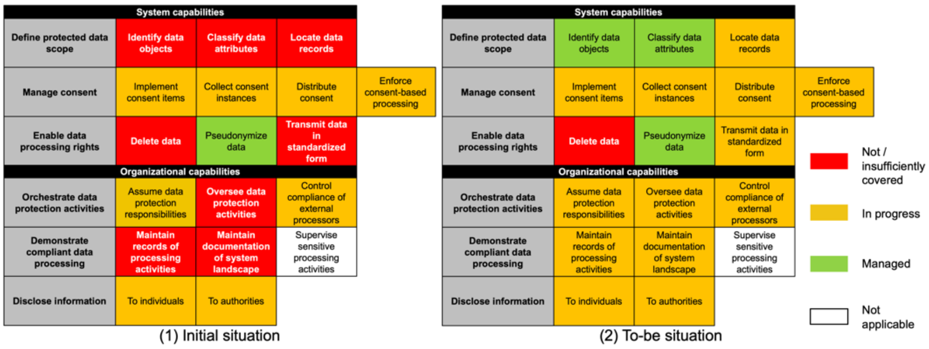


Figure 3. Evaluation of Leares' compliance level

Using the capability model contributed to alleviate these issues and helped Leares in identifying compliance gaps as well as defining and prioritizing actions. Going through the model, we were able to assign each check-list activity to capabilities, and assess to what extent they contributed to achieving compliance. When capabilities were partially covered by those activities, the model provided guidance to refine them. The capability model also helped identifying capabilities that Leares had not considered at all, such as defining the protected data scope. In these cases, new measures had to be defined. The instantiations depicted in Figure 3 show how the model was used to assess existing practices in the initial situation, along with a realistic target situation to be achieved within the next months. As a result of using the capability model, Leares was able to devise a structured action plan, covering all aspects of its data management practices.

5 Conclusion and Outlook

This paper introduces a data management perspective to EU-GDPR and argues that the regulation requires companies to build a dedicated data management capability. The suggested capability model was developed in an iterative design science process, integrating both interpretation of legal texts and practical insights from focus groups with more than 30 experts and from 3 EU-GDPR projects. By translating compliance requirements into organizational and system capabilities, it contributes to (1) building common ground between legal and data management domains and (2) assisting organizations in assessing practices, identifying and deciding on implementation options for achieving compliance with EU-GDPR. From a research perspective, our capability model complements the emerging body of research on EU-GDPR, that mostly investigates selected information privacy practices. The capability model may be used to classify and integrate these focused research efforts into an enterprise-wide perspective. Furthermore, it complements IS security research by focusing on non-security aspects of information privacy. For practice, the capability model supports companies in developing a systematic approach towards achieving EU-GDPR compliance and monitoring progress, instead of “fire-fighting”. As outlook for future research, our focus group discussions reveal that implementing EU-GDPR is not a one-time effort, but an ongoing process. The suggested capability model may serve as a basis for studying how the capabilities are being built and how they can be assessed. As it is supposed to contribute to reaching a firm’s control objective, a potential lead for further research would be to propose indicators of compliance goals and measure them.

Acknowledgements

This research was supported by the Competence Center Corporate Data Quality (CC CDQ). The authors would like to thank the experts that contributed to this research.

References

1. The Economist: The World’s Most Valuable Resource Is No Longer Oil, But Data, (2017).
2. De Hert, P., Papakonstantinou, V.: The Proposed Data Protection Regulation Replacing Directive 95/46/EC: A Sound System for the Protection of Individuals. Computer Law & Security Review. 28, 130–142 (2012).
3. The Race to GDPR: A Study of Companies in the United States & Europe. Ponemon Institute (2018).
4. Bélanger, F., Crossler, R.E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly. 35, 1017–1042 (2011).
5. Nicolaidou, I.L., Georgiades, C.: The GDPR: New Horizons. In: Synodinou, T.-E., Jougleux, P., Markou, C., and Prastitou, T. (eds.) EU Internet Law: Regulation and Enforcement. pp. 3–18. Springer International Publishing, Cham (2017).
6. Mitrou, L.: The General Data Protection Regulation: A Law for the Digital Age? In: Synodinou, T.-E., Jougleux, P., Markou, C., and Prastitou, T. (eds.) EU Internet Law: Regulation and Enforcement. pp. 19–57. Springer International Publishing, Cham (2017).
7. De Hert, P., Papakonstantinou, V.: The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals? Computer Law & Security Review. 32, 179–194 (2016).
8. Kurtz, C., Semmann, M., Böhmann, T.: Privacy by Design to Comply with GDPR: A Review on Third-Party Data Processors. In: AMCIS 2018 Proceedings (2018).
9. Petkov, P., Helfert, M.: Identifying Emerging Challenges for ICT industry in Ireland: Multiple Case Study Analysis of Data Privacy Breaches. In: AMCIS 2017 Proceedings (2017).
10. Karyda, M., Mitrou, L.: Data Breach Notification: Issues and Challenges for Security Management. In: MCIS 2016 Proceedings (2016).
11. Engels, B.: Data Portability and Online Platforms The Effects on Competition. In: BLED 2016 Proceedings. pp. 19–22 (2016).
12. Alboaie, L.: Towards a Smart Society through Personal Assistants Employing Executable Choreographies. In: ISD 2017 Proceedings (2017).
13. Fox, G., Tonge, C., Lynn, T., Mooney, J.: Communicating Compliance: Developing a GDPR Privacy Label. In: AMCIS 2018 Proceedings (2018).
14. Russell, K.D., O’Raghallaigh, P., O’Reilly, P., Hayes, J.: Digital Privacy GDPR: A Proposed Digital Transformation Framework. In: AMCIS 2018 Proceedings (2018).
15. El Kharbili, M.: Business Process Regulatory Compliance Management Solution Frameworks: A Comparative Evaluation. In: APCCM 2012 Proceedings. pp. 23–32 (2012).

16. Cleven, A., Winter, R.: Regulatory Compliance in Information Systems Research - Literature Analysis and Research Agenda. In: Enterprise, Business Process and Information Systems Modeling. pp. 174–186. Springer-Verlag, Berlin, Heidelberg (2009).
17. Abdullah, N.S., Indulska, M., Shazia, S.: A Study of Compliance Management in Information Systems Research. In: ECIS 2009 Proceedings. pp. 1–10 (2009).
18. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly. 28, 75–105 (2004).
19. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems. 24, 45–77 (2007).
20. Bensoussan, A., Avignon, C., Bensoussan-Brulé, V., Forster, F., Torres, C.: Règlement Européen sur la Protection des Données: Textes, Commentaires et Orientations Pratiques. Bruylant, Brussels (2018).
21. Debet, A., Massot, J., Métallinos, N.: Informatique et libertés: la protection des données à caractère personnel en droit français et européen. Lextenso, Issy-les-Moulineaux (2015).
22. Voigt, P., Von Dem Bussche, A.: The EU general data protection regulation (GDPR): A Practical Guide. Springer International Publishing, Cham (2017).
23. Guadamuz, A.: Developing a Right to be Forgotten. In: Synodinou, T.-E., Jougleux, P., Markou, C., and Prastitou, T. (eds.) EU Internet Law: Regulation and Enforcement. pp. 59–76. Springer International Publishing, Cham (2017).
24. European Data Protection Board: Guidelines On Consent Under Regulation 2016/679 (WP259, rev.01). EDPB (2018).
25. European Data Protection Board: Guidelines on Data Protection Officers (WP243 rev.01). EDPB (2017).
26. European Data Protection Board: Guidelines on Transparency under Regulation 2016/679 (WP260 rev.01). EDPB (2018).
27. Iannopollo, E., Balaouras, S., Harrison, P.: The Five Milestones to GDPR Success. Forrester Research (2017).
28. Merlivat, S., Iannopollo, E., Parrish, M., Khatibloo, F., Oesterreich, M., Liu, S., Turley, C.: Digital Advertising under GDPR Hinges on Data Management. Forrester Research (2017).
29. Peyret, H., Cullen, A., McKinnon, C., Blissent, J., Iannopollo, E., Kramer, A., Lynch, D.: Enhance your Data Governance to Meet New Privacy Mandates. Forrester Research (2017).
30. Iannopollo, E., Balaouras, S., Pikulik, E., Dostie, P.: The State of GDPR Readiness. Forrester Research (2018).
31. Deutsche Telekom: Binding Interpretations: General Data Protection Regulation (GDPR). Deutsche Telekom (2016).
32. Nickerson, R.C., Varshney, U., Muntermann, J.: A Method for Taxonomy Development and Its Application in Information Systems. European Journal of Information Systems. 22, 336–359 (2013).
33. Sadiq, S., Governatori, G., Namiri, K.: Modeling Control Objectives for Business Process Compliance. In: Alonso, G., Dadam, P., and Rosemann, M. (eds.) BPM 2007 Proceedings. pp. 149–164. Springer-Verlag, Berlin, Heidelberg (2007).
34. Zhang, M., Sarker, S., Sarker, S.: Drivers and Export Performance Impacts of IT Capability in 'Born-Global' Firms: a Cross-national Study. Information Systems Journal. 23, 419–443 (2013).
35. Grant, R.M.: The Resource-Based Theory of Competitive Advantage: Implications for Strategy Formulation. California Management Review. 33, 114 (1991).
36. Baiyere, A., Salmela, H.: Towards a Unified View of Information System (IS) Capability. In: PACIS 2014 Proceedings (2014).
37. Bharadwaj, A.: A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. MIS Quarterly. 24, 169–196 (2000).

On the Difficulties of Incentivizing Online Privacy through Transparency: A Qualitative Survey of the German Health Insurance Market

Max Maass[1], Nicolas Walter[1], Dominik Herrmann[2], and Matthias Hollick[1]

[1] TU Darmstadt, Secure Mobile Networking Lab, Darmstadt, Germany
{mmaass,nwalter,mhollick}@seemoo.tu-darmstadt.de
[2] University of Bamberg, Privacy and Security in Information Systems, Bamberg, Germany
dominik.herrmann@uni-bamberg.de

Abstract. Today, online privacy is the domain of regulatory measures and privacy-enhancing technologies. Transparency in the form of external and public assessments has been proposed for improving privacy and security because it exposes otherwise hidden deficiencies. Previous work has studied privacy attitudes and behavior of consumers. However, little is known on how organizations react to measures that employ public "naming and shaming" as an incentive for improvement. We performed the first study on this aspect by conducting a qualitative survey with 152 German health insurers. We scanned their websites with PrivacyScore.org to generate a public ranking and confronted the insurers with the results. We obtained a response rate of 27%. Responses ranged from positive feedback to legal threats. Only 12% of the sites – mostly non-responders – improved during our study. Our results show that insurers struggle due to unawareness, reluctance, and incapability, and demonstrate the general difficulties of transparency-based approaches.

Keywords: privacy, web security, transparency, quasi-experiment, e-mail survey

1 Introduction

Privacy plays an increasingly important role for consumers, regulators, and companies, especially on the internet. A growing number of companies provide services in the areas of online advertising, web analytics, and user profiling. Users and regulators have responded by deploying tracking blockers and passing stronger privacy laws, e.g., the European General Data Protection Regulation (GDPR).

So far, most research has focused on user perceptions of information privacy [1–4] and models for firms' data sharing gathering [5–7] and sharing behavior [8]. However, there is a lack of research on how to incentivize companies to reduce their use of tracking services, which are disliked by many users [9]. The study presented in this paper investigates the role of *transparency* as an incentive mechanism. Our scope is not limited to *privacy* (represented by online tracking services); we also consider the

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

connected area of *security* (e.g., transport encryption of web and mail traffic), as they affect user privacy against malicious actors like criminals and intelligence agencies.

Existing website scanners like the Qualys SSL Test (ssllabs.com) and Webbkoll [10] allow to assess security and privacy features of single sites. However, these scanners communicate the result only to the particular user who commissioned a scan. Therefore, there is little incentive for site operators to improve their rating. In contrast, we consider a scenario in which the scan results for multiple competing organizations are published openly on the internet in a ranking. This combination of publicity and comparability creates more transparency for consumers, which may increase the pressure on site operators to improve. Therefore, we pose the following research questions:

- **RQ1**: How do website operators react when they are notified that their website has been rated in terms of privacy and security aspects and the results are publicly available online?
- **RQ2**: Does telling them about being in a *public ranking* change their reaction?

We seek to answer these questions by contacting 152 German health insurance providers in a qualitative study, confronting them with and asking their comment on the current state of information privacy on their websites as determined by the PrivacyScore platform [11]. PrivacyScore.org is a public web service that automatically analyzes websites for security and privacy issues. Since its inception in June 2017, it has performed over 1 million scans and is being used by activists, data protection officers, and the general public. To the best of our knowledge, our study is the first to investigate how the competitive nature of public privacy and security rankings affects website operators, using the ranking functionality offered by PrivacyScore.

2 Related Work

The majority of IS privacy research focuses on individual privacy [6]. One of the central questions in this area is how privacy concerns affect the way individuals behave and their willingness to disclose personal information [1]. A frequently discussed aspect in this context refers to the decisions that individuals make regarding privacy and a trade-off between risk and benefit [3], which is informed by many factors, including company culture, legal environment, and the industry of the information-gathering company [4].

In contrast, there is less research on the question of how privacy issues are perceived from the **organization perspective**, i.e., by companies who work with personal data of their consumers. This field concerns itself with the interests and attitudes and motivations that companies have towards privacy [1, 5–8]. Greenaway and Chan argue that companies can behave in a reactive or proactive manner in regard to privacy and thus influence the perception of customers, and that their behavior can be explained using two models [5]. The *Institutional Approach* (IA) considers firms behavior as a search for legitimacy in the face of external pressures, and distinguishes an *acquiescent* (compliance with law, imitation of peers) and a *proactive approach* (exceeding minimum requirements and using clear communication to achieve leadership without falling out of line). The *Resource-based View* (RbV) posits that firms seek a sustainable competitive advantage using either an *information-* (through superior data analysis) or a *customer focus* (through superior customer trust).

To the best of our knowledge, the role of **competition** in promoting privacy has not received significant attention so far, although individual companies like Apple[1] or DuckDuckGo[2] have started marketing themselves as privacy champions. Like the RbV, Ohlhausen and Okuliar view privacy as a factor in competition [12]. Kerber reverses the argument and posits that a lack of competition may be partially responsible for the lack of privacy offerings from online companies [13].

Another reason may be that the privacy behavior of companies is often invisible from the outside. Thus, consumers find it hard to check whether companies are living up to their claims, for instance by following accepted best practices that ensure the privacy and security of their customers. **Transparency** can make these invisible practices of companies visible, enabling customers to easily verify claims and compare the practices of competing companies. The IS transparency literature has so far been focused on transparency strategies for companies [14], but has not yet considered transparency provided by external actors in the field of information privacy. Such external transparency has been shown to improve outcomes in areas such as Corporate Social Responsibility [15] and corporate ethics [16], among others, making it a promising avenue for more research.

3 Technical Background

We use the platform PrivacyScore.org [11] to investigate how operators of websites react when they are confronted with the fact that a third party is creating transparency about the privacy and security practices exhibited by their websites.

PrivacyScore.org[3] is an automated website scanner that allows anyone to investigate websites regarding privacy and security issues. Compared to other website scanners like Qualys SSL Scan and Webbkoll, the scans conducted by PrivacyScore are more comprehensive, including both security and privacy aspects. Currently, PrivacyScore performs more than 60 checks in four groups: *Tracking and Privacy* checks if a website includes third-party trackers. *Website Encryption* checks if a web server offers state-of-the-art secure connections, whether it redirects users to the secure version of the site, and whether it sets the HTTP Strict Transport Security (HSTS)[4] header, which prevents certain classes of attacks. *Mail Encryption* checks whether the mail servers of a website support state-of-the-art encryption, and *Web Security* checks if a website leaks internal information [17] and whether it protects users against client-side attacks.

Detailed results of each scan are published on the website and retained in a database to facilitate longitudinal studies. While it is possible to scan individual websites with

[1] https://www.cnbc.com/2018/10/25/in-privacy-fight-apple-has-more-to-lose-than-facebook-or-google.html (accessed 2018-11-16)

[2] https://www.theverge.com/2017/1/25/14381138/duckduckduckgo-privacy-anonymous-searching-10-billion (accessed 2018-11-16)

[3] The first and third authors of this paper are members of the PrivacyScore team.

[4] https://tools.ietf.org/html/rfc6797 (accessed 2018-11-19)

PrivacyScore, it is primarily intended to be used to scan lists of related websites (e.g., all pharmacies with an online shop in a specific country). Scanning a list generates a public ranking with all sites that belong to a list. Rankings put the individual scan results into context and allow consumers, data protection authorities, and site operators to assess how a particular site compares with its competitors. At the time of writing PrivacyScore hosts over 200 scan lists and has conducted more than 1 million scans.

A noteworthy limitation of PrivacyScore is that it conducts scans automatically without human oversight. As a result, the assessment is limited to aspects that can be reliably verified programmatically from the outside by visiting the website with an instrumented web browser. In particular, PrivacyScore cannot verify whether statements in the privacy policy are adequate and whether the site actually adheres to them. Moreover, PrivacyScore conducts only benign checks. For instance, it does not actively attempt to find and exploit vulnerabilities in the website under test.

Finally, PrivacyScore attempts to find a fair balance between satisfying the desire for transparency of consumers and operational interests of site operators. To this end, all scans are conducted instantaneously, i.e., without asking operators for permission. Website operators can have their website excluded from future scans. For reasons of transparency, this fact is published next to the scan results. Moreover, the last successful scan result before the complaint is shown for the respective site.

4 Study Design

As proposed by Greenaway and Chan [5], we use a mixed method [18] approach by combining a classical open question survey with a quasi-experimental setup based on the PrivacyScore platform. We follow a qualitative research process consisting of planning, data gathering, preparation of analysis, and analysis and summarization [19].

4.1 Gathering Websites and Scan Results

Our focus is on the health insurance sector within Germany, as visitors of health insurance websites may disclose sensitive information (e.g., by looking up information about specific illnesses and treatments). We obtained the homepage addresses (domain names) of all public and private health insurers in Germany from Wikipedia,[5] uploaded the URLs to PrivacyScore and added them to the newly created scan list “Deutsche Krankenkassen und Krankenversicherungen” (privacyscore.org/list/15). In December 2017 (T1) we saved a local copy of the scan result for each website as well as its rank within the aforementioned list. We performed two additional scans at later points in time (February and August 2018; T2 and T3) to determine whether the insurers had changed security and privacy properties of their sites.

[5] https://de.wikipedia.org/w/?title=Liste_deutscher_Krankenkassen&oldid=180861868 and https://de.wikipedia.org/w/?title=Liste_deutscher_privater_Krankenversicherer&oldid=1807 52925 (accessed 2018-09-11)

4.2 Group Allocation

Seeking to answer RQ2 we use a quasi-experimental setup, which is common in organizational research [20]. We split the insurers into two groups A and B, introducing the variable *competition*. We prepared two kinds of solicitation emails. Insurers in Group A were only informed about the overall scan results for their site in terms of the four areas that are published on the PrivacyScore website: "Tracking and Privacy", "Website Encryption", "Mail Encryption", and "Web Security". Insurers in Group B were additionally confronted with the rank of their site in our "Deutsche Krankenkassen und Krankenversicherungen" scan list. All solicitation emails started out with a brief description of PrivacyScore and ended with a request to participate in our study by answering the following two open questions: *What do you think about this kind of an assessment from a company's point of view? Would you consider making changes to your website in order to improve its privacy properties?*

The group allocation was not randomized but based on the ranking of the websites on PrivacyScore. This was done to ensure that both groups are homogenous in terms of the distribution of the ranks as given by the scan list. To this end, we assigned the numbers from 1 to 152 to the insurers in the order of the PrivacyScore ranking. The companies with odd numbers were put into Group A, those with even numbers into Group B. This procedure was guaranteed to result in two groups of the same size and allowed us to handle cases of ties in a deterministic fashion. Note that our allocation resulted in an uneven distribution in terms of the *type of insurer*: Group A contained 61 public and 15 private insurers, whereas Group B contained 51 and 25, respectively. This heterogeneity will be taken into account during the analysis.

4.3 Sending of Solicitation Mails

We obtained an email address for every insurer by visiting each website and manually searching for a suitable contact address in the privacy policy, the contact section, and the imprint of the site. If available, we preferred data protection-specific addresses such as "datenschutz@" to general-purpose addresses like "service@" or "info@".

On 8 December 2017, we sent out solicitation emails to all 152 insurers, using a university mail account not affiliated with the PrivacyScore project. As most insurers had not replied with a definitive response after four weeks, we sent out reminders to 135 insurers on 11 January 2018. The majority of responses was received in January and February 2018. After we had received all responses (March 2018) we iterated over them multiple times for open coding. As two of the authors are part of the PrivacyScore team, we were also in the position to receive any complaints directed towards the platform (which was not communicated to the recipients).

5 Results

In this section, we report the results of the initial scan of the websites and how site operators responded to our solicitation mails. Afterwards we discuss the three primary response types, i.e., positive responses, complaints, and other responses, in more detail.

5.1 Initial State of Websites

At the beginning of the study (T1), 26% (39) of the websites did not use any third-party tracking. 67% (102) of the websites used a well-configured Transport Layer Security (TLS) setup (defined as automatically forwarding users to the encrypted version of the website, offering TLS 1.2 and not offering the outdated and insecure protocol versions SSLv2 and SSLv3), and only 26% (39) of the websites set a HTTP Strict Transport Security (HSTS) header. Mail server TLS scans exhibited a failure rate of 6.5% (15), likely due to spam protection techniques like tarpitting. Still, 78% (119) websites had a well-configured mail server TLS setup (with TLS 1.2 and without SSLv2 and SSLv3).

5.2 Contacts and Respondents

In total, we received responses from 41 insurers (response rate: 27%). We found that sending a reminder message greatly increased the response rate, with more than half of the responses (23 of 41) reaching us only after the reminder. We had contacted 97 data protection (DP) contacts (64% of 152 insurers; A=45, B=52) and 55 general-purpose contacts (A=31, B=24). The overall response rates of Group A and B are similar (A=26%, B=28%), with data protection contacts showing a slightly higher rate (A=29%, B=31%) than general-purpose contacts (A=23%, B=21%).

We also observed many instances of messages being forwarded inside the companies, with different departments responding to our inquiries. An overview is given in Table 1. In Group A, our messages were frequently answered by marketing and data protection or IT teams (40% and 35%, respectively). Members of the board of directors (20%) and other departments (5%) sent the remaining 25% of responses.

In Group B, we observed a markedly different behavior. The majority of the responses were written by data protection and IT specialists (71%), while the remainder was evenly split between marketing, board members, and others. This deviation can only partially be explained by the larger number of DP contacts in Group B (A=45, B=52): The fraction of data protection and IT responses is higher in Group B for both messages sent directly to a DP contact (A=35%, B=75%) and for messages sent to a general-purpose contact (A=14%, B=60%).

5.3 Types of Responses

Overall, the responses fall into one of the following three categories: Firstly, there are *positive responses*, e.g., statements of gratefulness, expressions of interest in the study, and very detailed discussions of the scan results. Secondly, there are *complaints* criticizing the nature of the unsolicited scans and the publication of the scan results. Thirdly, there are *other responses*, ranging from acknowledgements of the receipt of our message to explicit expressions of indifference. The heat map in Fig. 1 provides a summary of the results by breaking down the responses according to their response type. It also indicates the relationship between response type and changes to the website, which we will analyze in Sect. 5.4 in more detail. All direct quotes in this paper are our translations of German replies.

Table 1. Recipients and Respondents

Group A			*Group B*		
Sent to	*Response From*	*Count*	*Sent to*	*Response From*	*Count*
General	DPO / IT	1	General	DPO / IT	4
	Marketing	3		Marketing	1
	Board Member	1			
	Other	2			
DPO	DPO / IT	6	DPO	DPO / IT	12
	Marketing	5		Marketing	1
	Board Member	1		Board Member	2
	Other	1		Other	1

Positive Responses. In total, we received 16 positive responses (A=9, B=7). In Group A, most responses came from the marketing (5) and data protection / IT departments (3), with a single response from a board member. In Group B, the responses came almost exclusively from data protection or IT departments (6), with only a single response from a marketing department. The responses varied widely in length and level of detail, ranging from single-sentence responses thanking us for the information and noting that the IT department would be investigating the report further, to differentiated technical and economic analyses of the trade-offs between user privacy and economic success for their company. For example, B_{58} explained that while they used tracking services on their websites, they were disabled in sensitive areas, and noted that "*tracking is above all about the care with which data is handled. The mere collection of data does not necessarily lead to better business success.*" A_{33} noted that while they did not forward users to the secure version of their website by default, all sensitive data (login information, payment details) were transmitted over a secure connection.

Two respondents claimed that they would be making changes based on the results of the scan: A_{67} gave a detailed response, referring to many individual test results in detail, and demonstrated that they had already reacted by implementing some changes, like enabling the Referrer-Policy[6] HTTP header. They also noted that some parameters related to the mail server were outside of their control, as they were managed by a third-party appliance. A_{63} responded that internal tests had confirmed the findings of PrivacyScore and mentioned nonspecific changes that were made to the website as a result. They also noted that "*the results helped us protect the privacy of the users of our website [and provided] valuable support for the implementation of changes based on an easier analysis and identification of weak points.*" In contrast, A_{35} noted the beta-status of PrivacyScore, but claimed that they would be performing their own checks to confirm the results, and act upon them, if necessary.

[6] https://www.w3.org/TR/referrer-policy/ (accessed 2018-11-19)

Response Type	Total	Insurer Type: Public	Insurer Type: Private	Group: Group A	Group: Group B	Website improved?: yes	Website improved?: no
Positive	11 %	7 %	3 %	6 %	5 %	1 %	10 %
Neutral	11 %	9 %	2 %	5 %	5 %	0 %	11 %
Complaint	6 %	6 %	0 %	2 %	4 %	0 %	1 %*
None	73 %	52 %	21 %	37 %	36 %	11 %	66 %
Sum	100 %**	74 %	26 %	50 %	50 %	12 %	88 %

*No scan results for 8 insurers that opted to be excluded from further scans. **Deviation due to rounding error.

Figure 1. Breakdown of responses according to response type

Finally, A_{91} provided a highly detailed response, going into some detail on the trade-offs between user privacy and economic success in online tracking. They noted that online tracking was required to evaluate the effectiveness of their online affiliate marketing campaigns. An additional concern was the analysis of the needs of potential customers, noting that "*we determine the needs of new and existing customers through market research studies. However, [...] their results are only of limited significance. This can be seen in the fact that statements determined by market research and actual user behavior partially contradict each other.*" Another concern was personalization, with the respondent noting that "*both market research and measurements show that personalized content [...] is used much more intensively by users and is increasingly expected, [which] can only be fulfilled by tracking tools and marketing automation. [...] Of course, users also have a high interest in sufficient protection of their privacy. However, we are convinced that the data protection regulations in force in Germany [...] cover the expectations of most users.*"

Several respondents noted that they were already using external scanners to validate the security of their websites and that they commissioned professional security audits for their websites. Almost all respondents (A=7, B=5) forwarded the report to their IT departments for further analysis. There were also cases of respondents not being aware of alternatives to the privacy-invasive technologies they were using, with one respondent showing surprise when being informed about privacy-enhancing alternatives to their current practices.

Complaints. We received nine complaints, all of them from public insurers. As Group A contains a higher ratio of public insurers, we would have expected more complaints from this group. However, the majority of complaints came from Group B (A=3, B=6). Most complaints came from companies whose site had a poor rank in the initial scan (min 25, mean 80, median 87, max 123 of 125). Four of the nine complaints (A=1, B=3) were raised directly with the PrivacyScore team, i.e., not by replying to our solicitation mail. In one case, we received a positive reply to our solicitation email, while the PrivacyScore team received a complaint two days later. Eight of the nine companies requested to be excluded from future scans.

In the following, we will give an overview of the concerns raised in the complaints. We observed that the tone of the responses was generally more terse in Group B. Five of the six complaints from Group B explicitly referenced the ranking. One of the three complaints from Group A explicitly referenced the fact that the results were publicly available. One frequently raised point was that the scan was performed (B=2) and published (A=1, B=2) without asking for permission in advance. Three companies (A=1, B=2) claimed to be investigating if the scan constituted an illegal attack on their infrastructure, B_{150} alleged a violation of competition law, and A_{109} argued that the scan violated their copyright. In contrast, legal scholars concluded that PrivacyScore is fully compliant with German law [21].

Of particular interest is the complaint by B_{150}, which we will discuss in more detail here. After obtaining a low rank in the initial scan, their information security officer contacted the PrivacyScore team, asking to exclude the website from future scans. Almost two months later another response reached the PrivacyScore team, this time from the chief legal officer, stating that the legal department of the company had analyzed the case and had raised a number of issues with how the results were being displayed. In particular, they objected to a perceived incomprehensibility on how the rankings were computed and how old results from excluded websites were maintained. They claimed that this could be "*damaging to their company*", and may be illegal under the law against unfair competition. At the same time, they acknowledged that they were open to critical analysis of their website. In a subsequent phone call, the company explicitly mentioned the competitive and privacy-sensitive nature of the public health insurance market and the disadvantages a company could experience from a low rank in such a privacy ranking. A constructive discussion resulted in several changes to PrivacyScore being proposed and implemented, adding various clarifications to the results pages to make it easier for visitors to understand what they (do not) imply. In return, the insurance company started work to improve some of the security and privacy properties of its website. However, they still declined to be added back to the ranking.

Other Responses. We received 16 neutral reactions (A=8, B=8). Six organizations (A=4, B=2) explicitly stated that they were not interested in participating in the study. Five others (A=1, B=4) replied that our message had been forwarded internally, but we did not receive another response from them. Organization A_{153} declined participation, citing insufficient capacity, and B_{40} stated that responding would entail a work order to an external service provider. A_{51} stated that they could not give any information on the topic of privacy. B_{94} states that they did not understand the provided information. Finally, A_{95} simply stated that they would not be making any changes to their website.

5.4 Changes to the Websites

To evaluate how websites changed over the course of the study, we performed measurements before sending our messages (T1) and repeated them at the end of the study (T2), skipping insurers that had asked to be excluded from future scans.

Twelve companies had stated that they were open to changing their website (A=6, B=6). However, at T2, only 4 (A=4) had actually made changes to one of the parameters measured by PrivacyScore. The total number of embedded third-party trackers (TPTs) increased over the course of the study period, from 411 to 439, which can be explained by a general inflation of the number of service providers in the market for behavioral advertising. We observe 12 (A=7, B=5) cases of at least one TPT being removed, and 13 (A=7, B=6) cases of at least one TPT being added.[7] Most of these websites belong to companies that did not respond to our messages. A_{63}, who had responded positively and promised to make changes, only replaced one TPT with a different one. A_{91}, who had provided us with a differentiated view of the trade-offs between privacy and economic success, appears to have briefly removed a number of third parties, but has since added them again. A_{133} had stated their willingness to adapt their website – and indeed, they have removed all trackers, leaving only a cookie consent script hosted by a third party. Conversely, A_{74} responded positively to our message, but only discussed the security aspects of the PrivacyScore evaluation, without regard to the privacy ratings. Their website added four additional trackers over the course of the study.

Several companies also made changes to the security of their systems by changing the configuration of their TLS setups. Four companies (A=3, B=1) disabled the outdated and insecure protocol version TLS 1.0, and two of them (A=1, B=1) also disabled TLS 1.1, leaving only the latest available version TLS 1.2 active. Five companies (A=3, B=2) enabled HTTP Strict Transport Security, while one (B=1) disabled it. However, none of them had responded to our message, so it is unknown if this happened as a reaction to our messages, or due to other, unrelated reasons. Similarly, 10 companies (A=6, B=4) started automatically forwarding all visitors to the secure version of their websites. Closer investigation revealed that while none of them had responded to our messages, five are maintained by the web design agency maintaining the website of B_{128}, who had asked to be excluded from future scans and threatened legal action. A manual visit of their website revealed that they, too, now forwarded all visitors to the secure version of their website. Thus, it is plausible that our messages were at least partially responsible for this agency-wide change.

Our observations indicate that insurers are willing to deploy changes that benefit the privacy interest of users without impacting the economic interests of the insurer (e.g., enabling a more recent TLS versions). If privacy and company interests are in conflict, companies are more reluctant to make a change (cf. the smaller number of insurers that removed third-party trackers from their site).

To put our results into context we compare the impact of our solicitation mails with the effect of regulatory changes (cf. Fig. 2). For this purpose, we scanned the websites once more in August 2018 (T3). Besides the introduction of the GDPR in May 2018, the timeframe also included new PCI DSS rules coming into effect in June 2018,[8] requiring websites that process credit card payments to update their TLS configuration.

Due to technical issues with PrivacyScore, scans for 10 sites reproducibly failed and had to be excluded in T3. Moreover, we observed significantly higher failure rates for the mail server TLS scans in T3 than in T1 and T2, precluding a meaningful comparison in this area. Cursory investigations indicate that the failures are due to the increasing prevalence of spam defense mechanisms deployed by mail server operators.

[7] We consider individual third parties, not the total number. Thus, if a website removes one third party and adds a different one, it will be counted in both groups.

[8] https://blog.pcisecuritystandards.org/are-you-ready-for-30-june-2018-sayin-goodbye-to-ssl-early-tls (accessed 2018-09-11)

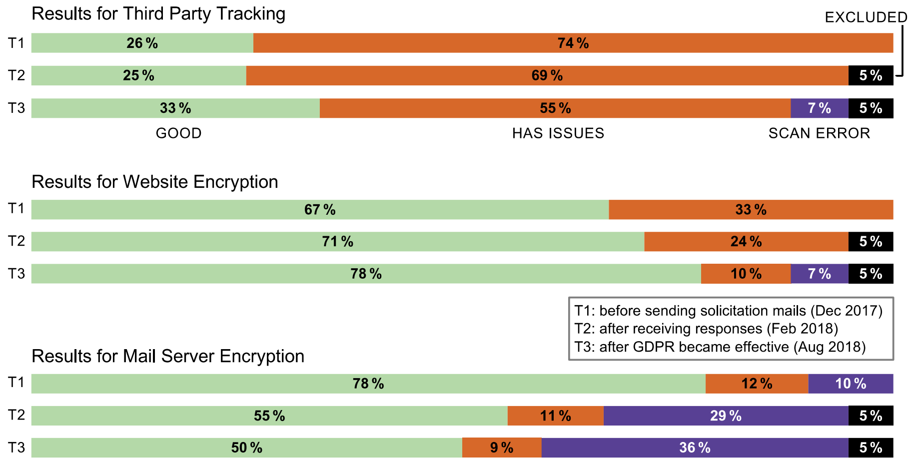


Figure 2. Scan results of health insurer websites at different points in time

As expected, the changes between T2 and T3 are substantial. 19 sites[9] removed all third parties, while 4 sites reintroduced trackers (compared to 2 and 1, respectively, in our study). 16 improved their TLS configuration (compared to 11 in our study), with 6 adding and 1 removing HSTS (5 additions and 1 removal in our study).[10]

Even though the time between T2 and T3 is much longer than between T1 and T2, it is unlikely that the observed differences are only due to the different durations given the changes to two regulatory frameworks. While we cannot infer what reasons served as an incentive for insurers to improve security and privacy features of their websites between T2 and T3, the combination of the passage of time and new regulatory requirements has a higher impact than our solicitation messages. This is to be expected, as the GDPR and PCI DSS changes were important and widely publicized, with strong incentives for compliance. While regulation has a large effect, transparency can still remain a valuable tool in affecting changes, as regulatory changes are infrequent events. Furthermore, regulation only establishes a lower bound of acceptable behavior, without incentives for exceeding the minimum requirements. Since the minimum requirements still permit many privacy-invasive techniques, transparency can serve as an incentive to exceed these requirements, which is vital for improving the state of online privacy.

6 Discussion

In this section, we discuss our results and their implications. The goal of our research was to investigate how health insurance companies react to transparency through public

[9] This includes 10 websites operated by the same association of insurers, whose websites are centrally managed. Thus, the number of distinct organizations making changes is at most 10.

[10] The difference with the changes visible in Figure 2 is due to some websites moving to the *excluded* or *failed* category

security and privacy ratings of their websites (RQ1) and if knowledge about the results being displayed in a ranking changes their response (RQ2).

We found that the responses varied greatly in the level of detail and content, ranging from detailed analyses of the trade-offs between user privacy and economic success to terse legal threats seeking removal from the public ranking. The responses cited a variety of reasons for the current state of their websites, ranging from conscious trade-offs between user privacy and the economic success of their company to technical limitations, e.g., third-party appliances whose configuration cannot be changed. One respondent also identified tension between the users' expectations of personalized content, which necessitates tracking, and user privacy.

6.1 Rankings Create Antagonism

We observed that public health insurers are more likely to complain than private insurance companies. Multiple complaints cited the privacy-sensitive nature of the health insurance market, and the potential competitive disadvantages caused by a bad privacy result. This led to a number of companies asking to be excluded from future scans, often under the threat of legal action based on questionable legal bases, including copyright, competition law, and cybercrime legislation [21].

We also observed a higher proportion of complaints from Group B (A=3, B=6), which was explicitly informed that their website was part of a ranking. This seems to indicate that the public ranking led to a higher probability of complaints being made. This theory is supported by the fact that five of the six complaints from Group B explicitly referenced the ranking in their complaints.

The complaints also have in common that all of them passed through the data protection and/or IT department of the companies, either as the initial point of contact for our study, or through company-internal forwarding. Thus, at first sight, a competing explanation for the higher ratio of complaints from Group B may be that the data protection and IT departments generally have a higher likelihood to complain. Coincidentally, we sent a higher fraction of our mails directly to data protection contacts in Group B (A=59%, B=68%), which corroborates the competing explanation. However, there are two arguments that challenge its validity: Firstly, several responders explicitly referenced the ranking in their complaints, and secondly, the complaint rate is still higher for Group B if only messages passing through data protection and/or IT departments are considered (A=21%, B=30%). Thus, we give more credence to our initial theory, which confirms RQ2.

6.2 Reasons for the Reluctance to Change

Our scans indicate that the websites under study are in permanent flux, with a general trend towards increasing the number of third-party trackers. As many of the insurers whose website changed during our study did not respond to our messages, the effect of our study on the websites is hard to quantify. Based on the received responses, we can attribute four changes directly to our messages, including one company that stopped using third-party trackers altogether, and assume that six more changes are at least partially the result of our messages. While these changes improve the privacy and security of website visitors, the low number of overall changes shows that the effectiveness of the messages was limited, indicating that at least in its current state, transparency and rankings through PrivacyScore do not significantly influence the willingness of most website operators to change their websites.

Many responses matched the *acquiescent approach* proposed in the IA [5], stating that they were in full compliance with applicable laws without giving additional details. A small number of companies moved towards the *proactive* approach by elaborating more on their internal processes, stating that tracking was disabled in critical areas, or citing frequent security audits of their websites.

We distil three reasons for the current state of insurance company websites: Conflicting value propositions, missing awareness, and negligence. Firstly, companies operate in a field of tension between their own economic goals (e.g., evaluating the effectiveness of their marketing campaigns), their reputation, and (sometimes conflicting) customer expectations like privacy and personalization. If not all expectations can be fulfilled, a trade-off needs to be found, leading companies to evaluate the costs of each solution. Stated in terms of the RbV [5], by gathering customer data (*intellectual resource*), companies can satisfy their own goals and some of the user expectations (e.g., personalization). As the use of web tracking is ubiquitous, it incurs almost no reputational cost and thus has low potential for differentiation (*relational resource*), leading most companies to pursue a *knowledge focus* and value their own economic goals higher than the privacy interests of their customers. This also explains their observed reluctance to be included in a ranking, making them favor an intentionally intransparent strategy, as proposed by Gerlach *et al.* [7].

Secondly, website operators may not be aware of alternative solutions that allow them to maintain the utility of their current solutions while decreasing their impact of the privacy of their users. Such alternative solutions include self-hosted tracking software like Matomo (matomo.org) that keeps the data under the control of the company, or two-click social media buttons (github.com/heiseonline/shariff) that do not disclose information to social networks on every page view.

Lastly, website operators may have been negligent and forgotten to configure their website correctly to respect the privacy of their users. This could manifest as the failure to enable the IP anonymization feature of Google Analytics (which is mandatory in Germany), or not forwarding visitors to the encrypted version of the website.

Problems caused by negligence can in many cases be remediated by a notification to the website operators, although such notifications have been shown to not always be reliable [17]. Raising awareness for privacy-preserving alternatives cannot easily be done at scale. In addition, awareness alone is not sufficient, as deciders need to be convinced that the benefits of switching to a privacy-preserving solution are worth the required effort and potential costs of changing the website, leading back to the issue of conflicting value propositions. These can only be influenced by changing the costs associated with the different options, which is easiest in the area of reputational costs. In the context of online privacy this means that privacy-invasive techniques have to become reputationally "expensive", which would turn forgoing their use into a relational resource, allowing differentiation through the *proactive approach*. This idea is at the center of the PrivacyScore project. However, our evaluation has shown that at the moment, the effect of transparency through PrivacyScore is not sufficient to cause large-scale changes. It is unclear if this is a general result or an artefact of the relatively unknown PrivacyScore platform. After all, companies may be more inclined to perform changes if a well-known consumer group or news outlet published a privacy ranking.

Finally, the behavior of companies can also be influenced by legislation and regulatory oversight. This was confirmed by our post-GDPR scan, which showed changes in a larger number of websites. Thus, the upcoming European ePrivacy regulation is a promising avenue for affecting further changes at scale.

7 Limitations, Future Research and Conclusion

Our study is subject to limitations. Firstly, we only investigate a single, privacy-sensitive sector – health insurance – in a single country. Extending and replicating the study with different sectors and in different countries could shed additional light on the general applicability of the results. Secondly, the number of respondents does not allow us to draw statistically significant conclusions. Our analysis also only considers two fixed dates when evaluating changes to the websites of the included companies, and we do not investigate if improvements are sustained over time or reverted.

Another limitation is the permeability of the group assignments. While we did not inform members of Group A about the ranking feature of PrivacyScore, the scan results page contains a (non-prominently presented) link to the ranking. Thus, members of Group A might have learned about it on their own. Nevertheless, we observed notable differences between Group A and B, which indicates that many members of Group A have not taken notice. The last limitation is the nature of PrivacyScore. As a relatively new platform, the publicity provided by it may present less of an incentive for change than a more popular and well-known platform or publication would have provided.

In conclusion, our results show that transparency – in the form of public assessments – can improve privacy features of websites. However, such efforts can also result in complaints and legal threats. A major factor limiting the willingness to change is the conflict between user privacy and the perceived need for privacy-invasive analytics for economic success: our solicitation mails led to much larger changes in areas where company and user interests are aligned, like website connection security. Another factor contributing to the current state appears to be a lack of awareness about privacy-preserving alternatives to common tracking services. While our study provided some initial insights on these difficulties, evaluating the effects of transparency on privacy remains a promising avenue for future work, for instance when publishing assessments in more widely disseminated channels like newspaper articles.

Acknowledgements. This work has been co-funded by the DFG as part of project C.1 within the RTG 2050 "Privacy and Trust for Mobile Users." The authors would like to thank Nora Wessels, Anne Laubach, Henning Pridöhl, and Pascal Wichmann for their assistance.

References

1. Bélanger, F., Crossler, R.E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Q. 35, 1017–1041 (2011).
2. Yun, H., Lee, G., Kim, D.J.: A Chronological Review of Empirical Research on Personal Information Privacy Concerns: An Analysis of Situational Contexts and Research Constructs. Inf. Manag. 1–13 (2018).
3. Smith, H.J., Dinev, T., Xu, H.: Information privacy research: an interdisciplinary review. MIS Q. 35, 989–1016 (2011).
4. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Inf. Syst. Res. 15, 336–355 (2004).
5. Greenaway, K.E., Chan, Y.E.: Theoretical Explanations for Firms' Information Privacy Behaviors. J. Assoc. Inf. Syst. 6, 171–198 (2005).
6. Greenaway, K.E., Chan, Y.E., Crossler, R.E.: Company information privacy orientation: a conceptual framework. Inf. Syst. J. 25, 579–606 (2015).
7. Gerlach, J.P., Eling, N., Wessels, N., Buxmann, P.: Flamingos on a slackline: Companies' challenges of balancing the competing demands of handling customer information and privacy. Inf. Syst. J. 1–28 (2018).
8. Conger, S., Pratt, J.H., Loch, K.D.: Personal information privacy and emerging technologies. Inf. Syst. J. 23, 401–417 (2013).
9. Plane, A.C., Redmiles, E.M., Mazurek, M.L., Tschantz, M.C.: Exploring User Perceptions of Discrimination in Online Targeted Advertising Exploring User Perceptions of Discrimination in Online Targeted Advertising. USENIX Secur. Symp. 19 (2017).
10. Andersdotter, A., Jensen-Urstad, A.: Evaluating Websites and Their Adherence to Data Protection Principles: Tools and Experiences. In: IFIP AICT vol. 498. pp. 39–51 (2016).
11. Maass, M., Wichmann, P., Pridöhl, H., Herrmann, D.: PrivacyScore: Improving privacy and security via crowd-sourced benchmarks of websites. LNCS. 10518, 178–191 (2017).
12. Ohlhausen, M.K., Okuliar, A.P.: Competition, Consumer Protection, and the Right [Approach] To Privacy. Antitrust Law J. 80, 121–156 (2015).
13. Kerber, W.: Digital Markets, Data, and Privacy: Competition Law, Consumer Law, and Data Protection. J. Intellect. Prop. Law Pract. 11, 856–866 (2016).
14. Parris, D.L., Dapko, J.L., Arnold, R.W., Arnold, D.: Exploring transparency: a new framework for responsible business management. Manag. Decis. 54, 222–247 (2016).
15. Lee, M., Kohler, J.: Benchmarking and transparency: Incentives for the pharmaceutical Industry's Corporate Social Responsibility. J. Bus. Ethics. 95, 641–658 (2010).
16. Maynard, M.L.: Policing transnational commerce: Global awareness in the margins of morality. J. Bus. Ethics. 30, 17–27 (2001).
17. Stock, B., Pellegrino, G., Li, F., Backes, M., Rossow, C.: Didn't You Hear Me? Towards More Successful Web Vulnerability Notifications. Proc. 2018 NDSS (2018).
18. Venkatesh, V., Brown, S.A., Bala, H.: Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. MIS Q. 37, 21–54 (2013).
19. Mayring, P.: Qualitative Content Analysis. Forum Qual. Sozialforsch. 1, (2000).
20. Cao, J., Crews, J.M., Lin, M., Deokar, A., Burgoon, J.K., Nunamaker Jr., J.F.: Interactions Between System Evaluation and Theory Testing: A Demonstration of the Power of a Mulitfaceted Approach to Information Systems Research. J. Manag. Inf. Syst. 22, 207–235 (2006).
21. Maass, M., Laubach, A., Herrmann, D.: PrivacyScore: Analyse von Webseiten auf Sicherheits- und Privatheitsprobleme. Informatik 2017. 1–13 (2017).

What is Your Selfie Worth? A Field Study on Individuals' Valuation of Personal Data

Nora Wessels[1], Amina Wagner[1], Jayesh Prakash Sarswat[1,2], and Peter Buxmann[1]

[1] Technische Universität Darmstadt, Software & Digital Business, Darmstadt, Germany
{wessels,wagner,buxmann}@is.tu-darmstadt.de
[2] jayesh_prakash.sarswat@stud.tu-darmstadt.de

Abstract. Referred to as the new oil, undoubtedly personal data is a valuable resource for organizations. Contrary, it is still blurred, to what extent individuals value their data even though, in a digitized world, users are requested to exchange their data for adequate services. Former research on individuals' valuation of personal data result in scattered, partly contradictious values, depending on the data type, context, and the measurement method. In this study, we aimed to facilitate the valuation for individuals by applying a new and promising measurement methodology: the participants of our field experiment had the chance to sell their selfies in a name-your-own-price auction with repeated bidding and feedback loops. As a result, 39% of our participants were willing to donate or sell their selfies with a median of 5€. Additionally, bidding clusters were identified. Implications for research on the valuation of personal data in terms of privacy are discussed.

Keywords: Value of Personal Data, Privacy, Willingness-to-Sell, Willingness-to-Accept, Name-Your-Own-Price Auction

1 Introduction

There is no question that we live in a period of time where personal information as a resource is becoming increasingly important. Discussions about personal data as the new oil, gold, or fuel are ubiquitous [1, 2] and the successes of data-driven companies like Google outcompetes traditional business models [3]. Notably, five of the top six companies according to their market value in the world are data-driven companies [4, 5]. Thus, personal information is a valuable asset for organizations and they already compete for profiting from it [6].

In contrast, taking the perspective of individuals, the benefits they get from releasing personal data is blurred as they typically do not receive a monetary compensation. Obviously, users benefit from using online services based on personal data, but still privacy concerns about the disclosure of personal information have increased, inherited with the feeling of being unfairly treated [7-9]. Indeed, a study in the Netherlands shows that individuals perceive that they are not getting enough value for their data, as 89% state that the industry benefits the most from data economy [10]. Thus, individuals seem to become more aware of the fact that their data is valuable, at least at an abstract level.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

However, to assess the current data handling processes and regulation approaches, individuals need to know the value of their data more precisely in order to be able to compare the costs of potential privacy losses in terms of data disclosure with the benefits of data usage [11]. In an attempt to investigate individuals' valuation of personal data, recent research measured either users' willingness-to-pay (WTP) for an enhanced privacy level or their willingness-to-sell (WTS) personal information [e.g., 12, 13, 14]. Yet, resulting valuations are scattered between studies depending on the type of data, the context in which the studies are conducted, and the research methodology [15]. So, it seems that individuals' valuation of personal data is not trivial.

Although personal data is reminiscent of traditional valuable resources, it is unlike these previous resources because of its infinity and the fact that it does not consume itself [16]. Further, personal data cannot be compared with ordinary goods for which individuals can set up a price more easily, as data can be sensitive and may make the disclosing person identifiable [17]. This makes it even more complicated for individuals to put a price tag on it. Therefore, research in the field of valuation of personal data requires carefully selected study designs and in turn measurement methods that can handle individuals' potentially vague awareness of the valuation of personal data due to its specificity and minor experience of its monetarization. In this vein, we have conducted a study to measure individuals' valuation of personal data with a new and promising methodology, which is already applied in other contexts for the pricing of 'opaque' goods [18]: a name-your-own-price (NYOP) auction. This methodology can identify the bidders' unique, lowest price for personal data separately for every individual, as it does not require synchronous bids [18-20]. Further, NYOP auctions can be implemented with a repeated bidding option that give individuals the chance to receive feedback if the bids are too high in order to state re-considered bids. This can facilitate individuals' establishment of the price, using the feedback as a source of information and as a trigger to reveal the lowest price [21, 22].

By strengthening the external validity of measuring individuals' value of personal information with a new methodology, we are aiming to answer the following research question: *What monetary value do users assign to their personal information, when being able to state several values after receiving feedback in a name-your-own-price auction with repeated bidding?*

In order to investigate this, we conducted a field experiment where participants could sell personal information in form of selfies in a name-your-own-price (NYOP) auction. We included a repeated bidding option, to make participants feel more comfortable with the situation as they could bid again after receiving feedback about the valuation and thus, can use the haggling process to overcome a lack of information due to low experience with monetized personal data [22]. Indeed, the results indicate that participants used their chances to bid several times, leading to a median final selling price of 5€. Analysis of the provided bids further show different clusters of bidders, indicating that the valuation of personal data is still very individual.

The paper is structured as follows: First, we summarize the theoretical background regarding the valuation of personal information in terms of privacy. In the subsequent chapter, the field study is presented with its methodology and results, before we close the paper with the implications, limitations, and future research directions.

2 Valuation of Personal Information

With the rise of data-driven business models, research and practice are concerned about the valuation of personal data from an individual perspective. Specifically, they are interested in individuals' valuation of privacy in order to understand their data disclosure decisions. In these attempts, scholars were either investigating individuals' willingness-to-sell (WTS) data [e.g., 23, 24] or their willingness-to-pay (WTP) for protecting their personal data [e.g., 12, 25] in order to measure the monetary value of data. However, former research studies are scattered and discordant [15]. Table 1 gives an overview of former studies investigating individuals' valuation of personal data along with their results, type of data under investigation as well as the applied measurement method.

Table 1. Overview of prior valuation of personal data studies
(foreign currencies are translated into € based on the exchange rate in November 2018)

Reference	WTS/ WTP	Type of Data	Research Method	Results (median or mean)
Danezis et al. (2005)	WTS	Location data	Vickrey Auction	11.42€ for research & 22.85€ for commercial purposes (median)
Cvrcek et al. (2006)	WTS	Location data	Vickrey Auction	43€ (median) for research & two-fold for commercial purpose
Brush et al. (2009)	WTS	Location data	Vickrey Auction	88.71€ for academic and commercial purposes (median)
Barak et al. (2012)	WTS	Location data	Close-ended questions	8€ (median)
Benndorf & Normann (2014)	WTS	Contact details & preferences on social networking sites (SNS)	Becker-DeGroot-Marschak method (BDM) & close-ended questions	**BDM** (mean): 8.32€ for preferences 14.88€ for contact details **Close-ended questions:** 24/24 accepted 5€ for preferences 35/42 accepted 5€ for contact details
Huberman et al. (2005)	WTS	Weight & age	Vickrey Auction	Age: 51.22€ (mean) Weight: 65.90€ (mean)
Grossklags & Acquisti (2007)	WTS & WTP	Weight	Surveys: close-ended & open-ended questions	**Open-ended Questions:** 12/12 accepted 0.22€ WTS-offers 12/14 accepted 0.89€ WTS-offers 1/7 accepted 0.22€ WTP-offers 1/14 accepted 0.89€ WTP-offers **Close-ended questions:** Min. WTS: 28.30€ (mean) Max. WTP: 0.71€ (mean)
Bauer et al. (2012)	WTP	SNS	BDM	0€ (mean)

Table 1. (continued) Overview of prior valuation of personal data studies

Reference	WTS/ WTP	Type of Data	Research Method	Results (median or mean)
Spiekermann & Korunovska (2017)	WTP & WTS	SNS	Contingent valuation method (CVM)	**WTP:** No market awareness: 0€ (median) Market awareness: 5€ (median) **WTS:** 49% WTS=0€ 23% 0€ < WTS < 4006€ 28% WTS > 4006€
Krasnova et al. (2009)	WTP	SNS	Conjoint Analysis	Between 14.14€ and 17.24€ a year
Schreiner & Hess (2015)	WTP	SNS	BDM	0.63€ (mean)

As shown in Table 1, the valuations for location data for instance vary between 11€ to 88€ even for the same measurement method of a reverse Vickrey auction due to different samples and variations in the study design [13, 26, 27]. By applying a survey with close-ended questions, Barak et al. identified a valuation of only 8€ [24] enlarging the range of results for location data even further. Supplementary, Benndorf & Normann found evidence that the measurement method can have a remarkable impact on the valuation, as the results for selling SNS details vary up to 10€ between Becker–DeGroot–Marschak mechanism (BDM) and close-ended questions for the same sample [28]. Similarly, Grossklags and Acquisti, ascertain that individuals requested a minimum price of 28€ for selling their weight information when being asked openly, while in the same study, individuals also accepted offers for even 0.22€ [14]. To spread the results even more, weight information was sold for a minimum WTS of 65€ in an auction study by Huberman et al. [23].

Grossklags and Acquisti also investigated individuals' willingness-to-pay for protecting their weight information and found a valuation of 0.71€ conforming a general gap between WTS and WTP [14]. The tendency that individuals demand more money for an object compared to the amount of money they are willing to pay for it, is widely known in research [29] and seems to hold for personal data as well [14]. Indeed, most WTP-studies report rather low values, however, also with variations within the prices. While Bauer et al. and Spiekermann & Korunovska for example found that individuals are not willing to pay a single Euro for their SNS details, the median increased to 5€ when data will be traded [6, 30]. Other studies, like the conjoint analysis by Krasnova et al. show that a user would be ready to pay 14€-17€ per year if no demographic information is used for personalized advertising, while Schreiner & Hess' BDM-participants reported to be willing to pay only 0.63€ for privacy control features within a Facebook premium version [12, 25].

To sum up, investigations on the valuation of personal information from an individual perspective is context-specific and depending on a variety of factors like the measurement method. The high discrepancies of individuals' valuation limit its implications for decision makers and theory. However, it also demonstrates that individuals have difficulties in assessing a stable and confident monetary value associated with personal information. This was also clearly confirmed by Brush et al. who reported that several of their participants found it challenging to value data and some even tried to ask for hints about what other participants had bid [27]. In our study we address the difficulties in assessing a value of personal information by relying on a novel measurement method. We chose a NYOP auction, because it facilitates individuals' valuation as they receive feedback. Further, it is not subject to a hypothetical bias as individuals were incentivized. In order to be transparent and unambiguous, we clearly explained how the data will be used and by which party.

3 Experimental Study

3.1 Methodology of the Experimental Study

In order to investigate individuals' valuation of personal data in an auspicious new way, we conducted a field experiment using a name-your-own-price auction with repeated bidding options where individuals could sell their selfies. In the following we will describe the experimental setting, the NYOP mechanism, and the study realization.

Experimental Setting. Following the call by Dinev et al. for more realistic study scenarios capturing actual behavior, we were aiming to provide an experimental design for our willingness-to-sell study, where participants perceive the disclosure of personal information as a natural and comprehensive task that leads to an actual sale [31]. As our natural environment is the university, it seems appropriate to develop a scenario that fits into this environment. Thus, students were the target group. We developed a fictional campaign, that our chair is looking for the "faces of our university" to advertise our institution among interested pupils and potential new students. Therefore, the alleged aim of this campaign was to collect selfies of students for marketing purposes in order to promote the university in an authentic and sympathetic way with the slogan "from students for students". For the purpose of this campaign, we implemented a website containing a detailed description of the presented purpose as well as the NYOP auction, with which the students could allegedly sell their selfies to us. We decided to focus on selfies as personal information, as selfies always depict the subjects' faces which turns a picture into personal information compared to ordinary photos. Furthermore, selfies are fashionable and omnipresent and thus, assures familiarity of young students and reduces the risk of misunderstanding [32].

Name-Your-Own-Price (NYOP) Auction with Repeated Bidding. This type of auction is based on an interactive pricing process between buyer and seller, in which both parties are actively involved in finding the price [33, 34]. It is referred to as a haggling process where no average or market price is disclosed [18, 35]. Traditionally, NYOP auctions are initiated by the product sellers to investigate buyers' willingness-to-pay without disclosing the lowest price [e.g., 35]. However, likewise studies building on other auctions [e.g., 13, 23, 26], NYOP auctions are also applicable to investigate individuals' willingness-to-sell. In doing so, the buyer sets a maximum price (also referred to as the treshold) at which he or she is willing to buy the good, but does not provide any information about it [19, 36]. The potential seller can then initate the first offer and if it meets or undershots the threshold, the sale is made in the amount of the value offered by the seller [19, 37]. However, if the bid is above the limit, it is rejected, but the seller has again the chance to repeat his or her bidding [18].

We opted for a NYOP auction for the following reasons: First, it is applicable for 'opaque' goods where the price is non-transparent [18]. Second, NYOP is used to elicit the lowest price of the bidder and thus, reflects individuals' actual willingness-to-sell transferrable to real-life situations [18]. Besides, in contrast to a "one price for all" strategy, it identifies individuals' unique selling price [19]. Fourth, depending on the design adjustments of the auction, the initiator of the auction can allow several bids in chronological order until the threshold is undershot (repeated bidding) [38]. Thus, individuals receive feedback if the bids are too high, which can serve as valuable information within the haggling process [22, 39], and therefore makes the method appropriate for individuals' vague awareness of the valuation of personal information. Indeed, former research has shown, that individuals' loose statement of a price without any reference information is associated with cognitive effort and therefore they prefer an alternative in which they can select a price or receive information about the valuation, for example by given reference price ranges [35]. However, this would go along with an impeding of the accuracy of the valuation [35] and a biasing in the direction of the external provided reference information [40]. In contrast, the feedback provided by a NYOP auction with repeated bidding option gives individuals something to go on in a subtler way without anchoring them too much as no initial value which then needs to be adjusted [41] is given. A fifth reason is that compared to reverse Vickrey auction often applied in studies investigating individuals' valuation of personal data [e.g., 23, 42], NYOP auctions do not only award one bidder [37]. So not only one participant has the chance to sell his or her personal information, but all who bid accordingly. Indeed, with NYOP there is no need to receive the bids at the same time, as asynchronously arriving bids can be accepted or rejected immediately if the threshold is set before [20].

Due to the design decision on repeated bids, we had to fix the threshold at a low value, because a higher value would have entailed the risk of losing information about a low willingness-to-sell. Thus, we decided to set the threshold to 1€.

We further decided to limit the amount of bids to three, as previous studies on the design of NYOP auctions are based on the assumption that sellers include frictional costs in their choice of bids [38]. Frictional costs are referred to mental efforts for navigating through the website, typing in the selected bids, as well as occuring from the waiting time until the bid is accepted or rejected [38, 39]. In order to prevent these frictional costs from taking on too much importance, we limited the number of bids to three. This also has the advantage of better comparability of the results of all participants in our study. Figure 1 summarizes the mechanism of our NYOP auction.

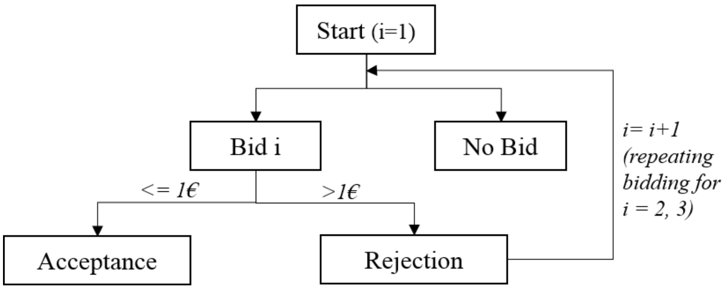


Figure 1. NYOP mechanism of our field experiment, adapted from [38]

Study Realization. We implemented a website for our NYOP auction. When entering the website, the participants were informed about our proposal of buying selfies for the chair's campaign, but we did not provide any price information. In addition to the NYOP mechanism, participants also had the opportunity to donate their selfies for the same presented purpose. This was even possible after making an offer. To meet the challenge that interested sellers might not have an appropriate selfie ready, we provided the chance to bid immediately, but upload the selfie later. However, we clearly stated that the money will only be paid off when the selfie is uploaded. For this purpose, we provided a separate link.

In order to distribute the website to students in our university, the field experiments' website was integrated into a cover online survey. Thus, the participants did not know the real intention of the study. For our cover study, we decided to emphasize on advertisement as a topic as it is omnipresent and therefore interesting for students. To provide students with an additional incentive to participate in the survey, they received 3€ in cash for responding to the online cover survey. The cover study consisted of different parts: After an introductory page providing all GDPR-relevant information, age and gender demographics were collected. Subsequently, different scales investigating participants' perception of advertising were presented. After finishing the cover study, we informed the participants about our current campaign where we are looking for the 'faces of our university' and that they have the chance to gain additional money by participating. By clicking on the next-button, all participants of the former cover study were forwarded to our NYOP website where a detailed description of the auction was given and interested participants could directly type in their bids or donate their selfies as presented above. The participants who were not interested in bidding or donating a selfie at all, were able to return to the initial cover survey by clicking on a "no interest"-button where we informed about the real purpose of the fictive campaign.

We advertised the cover study by distributing flyers on the campus and by setting up a post on our chairs' Facebook profile. From the 186 responses, we had to remove 15 due to bad data quality (incomplete or invalid data), leading to 171 complete data sets. The final sample consisted of 44 females (26%), 125 males (73%), and two unspecified genders (1%). The male quota is representing the gender distribution of our technology-focused university. With a share of 75%, the majority of the participants were between 18 and 25 years as expected due to our targeting on students. Further, 22% of the participants were between 26 and 35, 2% were between 36 and 45, and 1% was older than 46 years. From the 171 participants, 67 individuals bid or donated, while the other 104 directly returned to the cover survey.

To analyze the data and present the distribution of the bids, we followed former studies on valuation of personal data applying reverse Vickrey auctions [11, 13, 43]. Additionally, we conducted a Wilcoxon signed-rank test and cluster analyses.

3.2 Results of the Experimental Study

The results of our field experiment are presented in the following sections, divided into insights about participants' willingness-to-sell selfies for the advertised campaign and their willingness-to-donate selfies.

Willingness-to-Sell Selfies. In total, 54 participants of our study (32%) bid. The bidders were on average 23.76 years old (SD = 3.77) and consisted of 20% women, 78% men, and 2% unspecified genders. While 40 participants made full use of the possibility to bid three times, seven individuals provided only two offers, and other seven stated just one value. Following the assumption that the last bid provided in a NYOP auction bid series represents the participant's actual willingness-to-sell, the final bid is of most interest. We further assume, that those who bid two values, have already reached their lowest, unique willingness-to-sell within the second offered value, thus, this is their last bid. Analogous, those who eventually provided only one value seem to have such a clear valuation in mind that they only want to express their lowest, unique value in one bid. Thus, we have allocated all of these bids under *final bids*. The mean of these final bids is 29.24€, with bids ranging from 997€ to 1 Cent. The median of these final bids is 5€. The big difference between mean and median and the relatively high standard deviation of 135.24 indicates that there are some very high valuations shifting the mean, which will be analyzed later on. In total, the threshold of 1€ was met or undershot 14 times leading to successful "sales" of selfies. Table 2 summarizes the statistics of the final bids together with the first and second bid.

Table 2. Statistics of the starting bids, second bids, and final bids (in €)

	First / Starting bid	Second bid	Third / Final bid
Mean (SD)	93.28 (221.91)	50.95 (154.93)	29.24 (135.24)
Median	25	10	5
N	40	47	54
[Max; Min]	[999.99; 2]	[998; 1.1]	[997;0.01]

Looking at the first and second bids, it is striking that the participants successively decreased the values of each bid. For analyzing the differences between the three bids, we conducted pairwise Wilcoxon signed-rank tests as our data was not normally distributed. This test is equivalent to a one-sample t-test conducted at signed ranks substituting the differences and is used for comparing the equality of medians of two samples [44, 45]. We compared the first bids with the second and the final bids as well as the second bids with the final bids. Since we were interested in differences between two bids, we had to exclude seven values from the test which only provided one bid. The participants who provided two bids are analyzed within the comparison of the second and final bids. The test statistics summarized in Table 3 show that all bids are significantly different from each other with a large effect size (r) based on Cohen's indexes [46, 47]. Based on the medians reported in Table 2, the first bids' median is 25€, while the second bid decreases by 40% leading to a median of 10€. The decrease from the second to the final bid is even higher with 50%.

Table 3. Wilcoxon signed-rank test statistics

	z	p-value	N	r
First bid – Second bid	-5.514	.000	40	.87
Second bid – Final bid	-5.912	.000	47	.86
First bid – Final bid	-5.513	.000	40	.87

We further investigated the distribution of the bid series in more detail. As some of the bid series were appreciably higher than others, we conducted hierarchical cluster analyses (with median clustering, single linkage, and ward methods) for testing whether there are different bidder groups [48, 49]. For these analyses we imputed the missing values in the bid series where only one or two values were named by filling the absent bids with their only stated (or in the case where only one bid is missing with the second stated) bid. This is again in line with our assumption that the participants specified their valuation within their *final bid* which was for seven participants the only stated value.

Surprisingly, the analyses revealed no clear clustering as the very high values were so scattered that it would lead to clusters with very few values. In a second step, we excluded the nine very high values from further analyses that were identified by the single linkage cluster analysis in addition to a z-based outlier-analysis. However, these bidders should not be seen as outliers in a traditional sense, as these are no measurement errors, but show the respondents' extraordinary high value proposition of personal data in terms of privacy. These might be therefore seen as a group of *privacy protectors*.

A re-iteration of the cluster analyses could identify three different clusters. Figure 2 depicts the scatterplot of these bid series as well as the allocation of them to the clusters.

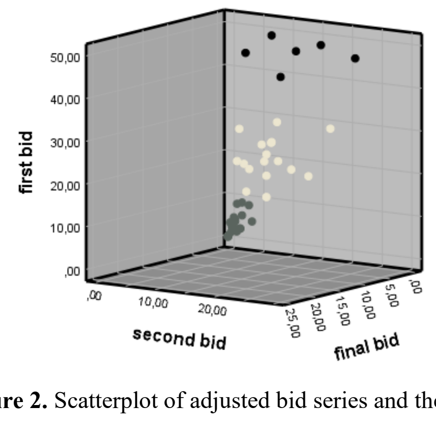


Figure 2. Scatterplot of adjusted bid series and their clusters

The first cluster, colored in black (Figure 2) is characterized by the fact that all bidders start with 50€ as first bids. The second bid is around 20€ and the final one about 10€ as depicted in Table 4. We call this cluster *striding out bidders* as the steps between the bids are with absolute decreases of 30€ and 10€ wider compared to those from the subsequent clusters. While the bidders in the white cluster show a similar final bid as the black cluster, their starting bid was only 25€ in median. Hence, their absolute steps between the bids were smaller compared with the former cluster. Thus, we name them *moderate bidders*. Finally, the bidders allocated to the grey cluster bid relatively low in all three bids. The median of the final bid in this *low bidders'* cluster meets our threshold of 1€.

Table 4. Statistics of the hierarchical cluster analysis (ward method) in €

Cluster (n)		First bid	Second bid	Final bid
Black (n=6)	**Mean**	50	20.83	11.20
Striding out bidders	**Median**	50	20	10
White (n=17)	**Mean**	24.16	15.41	10.01
Moderate bidders	**Median**	25	15	10
Grey (n=22)	**Mean**	4.36	2.91	1.84
Low bidders	**Median**	4.25	2	1

Table 5 summarizes the demographics of the three bidder groups identified by the clusters analyses as well as the *privacy protectors*. However, an ANOVA test could not identify significant statistical differences between the groups with regard to age or gender of the participants.

Table 5. Demographics of the bidder groups

Bidder group	Mean age (in years)	Women	Men	Others
Privacy protectors	25.44 (SD = 4.79)	11.11%	77.78%	11.11%
Striding out bidders	21.17 (SD = 1.34)	16.67%	83.33%	0%
Moderate bidders	23.71 (SD = 3.59)	17.65%	82.35%	0%
Low bidders	23.82 (SD = 3.46)	27.27%	72.73%	0%

Willingness-to-Donate Selfies. Additionally, participants had the chance to donate their selfies for the stated purpose. In total, 13 participants chose this option. From these, 38% were female, 62% male. The average age was 24.69 years (SD = 3.07). Three of these donators started to bid values in advance: One person bid 15€ before deciding to donate, another individual tried a 3€-bid first, and the third person bid 3€ and even 1€ so that the offer would be accepted if not donated. All results are discussed and implications are provided in the following.

4 Discussion

Within this study, we investigated individuals' valuation of personal data by conducting a NYOP auction with repeated bidding and feedback loops. In the following the implications as well as limitations and future research suggestions are discussed and a conclusion is given.

4.1 Implications of the Study

Our study adds to research in several ways. First, our main *theoretical contribution* is the investigation of the value individuals assign to their personal information in a realistic, highly accessible, and comprehensive fashion. In this vein, we relied on a new and promising method, a NYOP auction with repeated bidding revealing individuals' lowest, unique selling price [18, 19]. Due to the opportunity to bid several times after receiving feedback, individuals feel more comfortable by stating their valuation as they can use the haggling process as a source of information [22] and re-think their valuation with each bid. However, bidders are not anchored as no initial price as a starting point is presented [41]. In contrast to other auctions, sellers in NYOP auctions with a repeated bidding option can strategically start with a higher, internal generated value and then sequentially approach their own lowest bid based on feedback [39], revealing individuals' actual WTS.

Further analyses reveal subsequent implications. Thereby, it is striking that except for seven participants stating one value, most of the bidders took the chance to bid twice or three times. The three bids were significantly different from each other, decreasing by 40% from the first to the second and 50% from the second to the final bid. Looking deeper into the distribution of the bid series itself, we can observe some scattered, high bid series, referred to as the group of *privacy protectors* and three clusters of bidders with similar valuations. We could see very low bids of 1€ in median within the *low bidders'* cluster. These make up 41% of the bid series. Another two groups of bid series, the *moderate bidders* (31% of the bidders) and the *striding out bidders* (11%), end up both at a final bid of about 10€ in median, but differentiate by their bidding strategies: While the bid series in the first cluster start with a higher bid of 50€, the bid series in the second cluster are more temperate, starting with 25€ in median.

Lastly, the results show that 39% of the participants were willing to sell or donate a selfie. The 13 donations indicate that these participants seem to believe in the good cause of our campaign and a monetary compensation is not required. We further hypothesize that the three individuals that first bid and then decided to donate, perceived the threshold as "not worth" and then rather wanted to support the campaign by a donation.

Beyond that, we show that our study participants would sell their selfie for 5€ in median. This monetary value seems convincing as current photo-selling platforms offer about 5€ to their users for one picture. For example, *Candidly Images* and *Foap* provide marketplaces where normal hobby photographers can sell their photos to interested parties [50, 51]. The photos are sold for around 10$ whereby the seller receives 5$

(about 4.39 € based on exchange rate in August 2018) as a reward. Thus, it is indeed a price where demand and offer come about.

To conclude, we found participants' willingness-to-sell personal data in this field study to be realistic, but due to the scattered high values and the three different clusters, still to be very individual as well. This implies that valuation of personal information in terms of privacy is very sensitive, context-specific, and individual. As a result, it is not generalizable across humans.

Beyond theoretical implications, our research also provides *practical contributions.* Not only since the German Chancellor Angela Merkel proposed to tax the sale of personal data, the investigation of individuals' valuation of personal data is becoming an important issue [52]. Already in 2014 Jentzsch indicated that a truthful valuation of personal data is an important matter in this time, as an increasing number of online platforms allow the sale of personal data to companies [42]. Indeed, more and more platforms appear (e.g., Datacoup, Data Fairplay, Datum) on which users can actively sell their personal data they are willing to disclose to interested companies and are therefore monetary compensated [53]. In this vein, our research contributes to practice by providing a mechanism that fosters the measurement of individuals' unique, lowest valuation of personal data which can also help providers of data-selling platforms to evaluate their prices and assess their business chances.

Secondly, when individuals use social networking sites (SNS) like Facebook or Instagram, they are agreeing to the term that the provider receives the license to use all shared images [54, 55]. Thus, individuals are using an apparent free service but they are paying with their personal information in terms of their photos and user behavior in general. In this vein, one could argue that SNS users do not value their photos as they give it up for free. However, our results provide evidence that Internet users indeed value their photos beyond free services in return.

4.2 Conclusion, Limitations, and Future Research Suggestions

Personal information has its value for organizations and users, at least at an abstract level. Whereby organizations trade personal information like an asset, individuals take a rather passive role in this process. In order to be able to assess current data handling approaches, individuals are requested to assign a value to their personal data. However, former research investigating users' valuation of personal data is scattered. Against this background, we conducted an experimental study with a NYOP auction allowing repeated bids and therefore feedback loops aiming to facilitate the valuation. In total, 39% of our participants were willing to sell or donate their selfies to the university for advertising purposes with a median price of 5€. The applied method can reveal individuals' actual, lowest, and unique valuation and can therefore help Internet users to pave the way from a 'passive spectator' to an 'active beneficiary'.

However, our study comes not without limitations. First, the sample of our experimental study consisted of more male than female participants due to the gender distribution of our technology-focused university. Although we saw in the data a tendency that women were more willing-to-donate, this tendency was not significant. Further, we relied on a German sample which could limit the generalizability of our study results for other cultures. Thus, we call for a deeper investigation of gender differences with a more balanced, international sample.

Further, as shown by the literature review, studies on the valuation of personal information are very context sensitive, meaning that the willingness-to-sell personal information and the actual requested value depends strongly on the circumstances under which the study was conducted [15]. Thus, we were aiming to optimize these circumstances by stating in an easy understandable manner who is buying, what data, for which purpose. Further, in our scenario, there were no complex partner structures involved, as the university was the only buying party which seems applicable as several willingness-to-sell studies were already successfully providing an experimental setup in the university environment [e.g., 13, 24, 26]. However, it might be interesting to conduct a NYOP auction with a repeated bidding option in another context in order to improve the transferability of our results.

Acknowledgements. This work has been co-funded by the DFG as part of project A.3 within the RTG 2050 “Privacy and Trust for Mobile Users".

References

1. Medium, https://medium.com/datareum/data-is-the-new-gold-e6eb1aeeb640 (Accessed: 11.08.2018)
2. Economist, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (Accessed: 11.08.2018)
3. Brynjolfsson, E., Hitt, L.M., Kim, H.H.: Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? SSRN (2011)
4. Statista, https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/ (Accessed: 25.07.2018)
5. Slotin, J., http://www.data4sdgs.org/news/what-do-we-know-about-value-data (Accessed: 25.07.2018)
6. Bauer, C., Korunovska, J., Spiekermann, S.: On the Value of Information - What Facebook Users Are Willing to Pay. In: European Conference on Information Systems Proceedings (2012)
7. Acquisti, A., Taylor, C., Wagman, L.: The Economics of Privacy. Journal of Economic Literature 54, 442-492 (2016)
8. Culnan, M.J., Armstrong, P.K.: Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. Organization Science 10, 104-115 (1999)
9. Culnan, M.J., Bies, R.J.: Consumer Privacy: Balancing Economic and Justice Considerations. Journal of Social Issues 59, 323-342 (2003)
10. DDMA, https://ddma.nl/download/53179/ (Accessed: 25.07.2018)
11. Carrascal, J.P., Riederer, C., Erramilli, V., Cherubini, M., de Oliveira, R.: Your Browsing Behavior for a Big Mac: Economics of Personal Information Online. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 189-200. ACM (2013)
12. Krasnova, H., Hildebrand, T., Guenther, O.: Investigating the Value of Privacy on Online Social Networks: Conjoint Analysis. In: International Conference on Information Systems Proceedings. Paper 173 (2009)
13. Danezis, G., Lewis, S., Anderson, R.J.: How Much is Location Privacy Worth? In: WEIS vol. 5, (2005)

14. Grossklags, J., Acquisti, A.: When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information. In: WEIS, (2007)
15. Wagner, A., Wessels, N., Buxmann, P., Krasnova, H.: Putting a Price Tag on Personal Information - A Literature Review. In: Hawaii International Conference on System Sciences (HICSS), vol. 51, Waikoloa Village, Hawaii (2018)
16. Bharosa, N., Luitjens, S., van Wijk, R., Pardo, T.: Panel: Removing the Barriers for Personal Data Management. In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 125. ACM (2018)
17. Spiekermann, S., Böhme, R., Acquisti, A., Hui, K.-L.: Personal Data Markets. Electronic Markets 25, 91-93 (2015)
18. Terwiesch, C., Savin, S., Hann, I.-H.: Online Haggling at a Name-Your-Own-Price Retailer: Theory and Application. Management Science 51, 339-351 (2005)
19. Hinz, O., Hann, I.-H., Spann, M.: Price Discrimination in E-Commerce? An Examination of Dynamic Pricing in Name-Your-Own Price Markets. MIS Quarterly 81-98 (2011)
20. Fay, S.: Partial-Repeat-Bidding in the Name-Your-Own-Price Channel. Marketing Science 23, 407-418 (2004)
21. Klemperer, P.: Auctions: Theory and Practice, Economics Group, Nuffield College, University of Oxford. Economics Papers (2004)
22. Liu, J., Dai, R., Wei, X.D., Li, Y.: Information Revelation and Customer Decision-Making Process of Repeat-Bidding Name-Your-Own-Price Auction. Decision Support Systems 90, 46-55 (2016)
23. Huberman, B.A., Adar, E., Fine, L.R.: Valuating Privacy. IEEE Security & Privacy 3, 22-25 (2005)
24. Barak, O., Cohen, G., Gazit, A., Toch, E.: The Price is Right?: Economic Value of Location Sharing. In: Proceedings of the 2nd ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, pp. 891-900. ACM, Zurich, Switzerland (2013)
25. Schreiner, M., Hess, T.: Why Are Consumers Willing to Pay for Privacy? An Application of the Privacy-freemium Model to Media Companies. In: European Conference on Information Systems Proceedings, (2015)
26. Cvrcek, D., Kumpost, M., Matyas, V., Danezis, G.: A Study on the Value of Location Privacy. In: Proceedings of the 5th ACM Workshop on Privacy in Electronic Society, pp. 109-118. ACM (2006)
27. Brush, A., Krumm, J., Scott, J.: Exploring End User Preferences for Location Obfuscation, Location-Based Services, and the Value of Location. Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 95-104, Copenhagen, Denmark (2010)
28. Benndorf, V., Normann, H.T.: The Willingness to Sell Personal Data. DICE Discussion Paper, No. 143., (2014)
29. Horowitz, J.K., McConnell, K.E.: A Review of WTA/WTP Studies. Journal of Environmental Economics and Management 44, 426-447 (2002)
30. Spiekermann, S., Korunovska, J.: Towards a Value Theory for Personal Data. Journal of Information Technology 32, 62-84 (2017)
31. Dinev, T., McConnell, A.R., Smith, H.J.: Research Commentary—Informing Privacy Research Through Information Systems, Psychology, and Behavioral Economics: Thinking Outside the "APCO" Box. Information Systems Research 26, 639-655 (2015)
32. Statista, https://www.statista.com/statistics/683924/us-adults-familiar-selfie-age/ (Accessed: 27.08.2018)
33. Spann, M., Bernhardt, M., Häubl, G., Skiera, B.: It's All in How You Ask: Effects of Price Elicitation Format on Bidding Behavior in Reverse-Pricing Markets. In: Proceedings of the

26th Annual Conference of the Society for Judgment and Decision Making (SJDM) Toronto, Canada, (2005)
34. Hann, I.-H., Hinz, O., Spann, M.: Dynamic Pricing in Name-Your-Own-Price Channels: Bidding Behavior, Seller Profit and Price Acceptance. In: Workshop on Information Systems and Economics (WISE), (2006)
35. Chernev, A.: Reverse Pricing and Online Price Elicitation Strategies in Consumer Choice. Journal of Consumer Psychology 13, 51-62 (2003)
36. Amaldoss, W., Jain, S.: Joint Bidding in the Name-Your-Own-Price Channel: A Strategic Analysis. Management Science 54, 1685-1699 (2008)
37. Spann, M., Tellis, G.J.: Does the Internet Promote Better Consumer Decisions? The Case of Name-Your-Own-Price Auctions. Journal of Marketing 70, 65-78 (2006)
38. Spann, M., Skiera, B., Schäfers, B.: Measuring Individual Frictional Costs and Willingness-to-Pay via Name-Your-Own-Price Mechanisms. Journal of Interactive Marketing 18, 22-36 (2004)
39. Hann, I.-H., Terwiesch, C.: Measuring the Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel. Management Science 49, 1563-1579 (2003)
40. Johnson, J.W., Cui, A.P.: To Influence or not to Influence: External Reference Price Strategies in Pay-What-You-Want Pricing. Journal of Business Research 66, 275-281 (2013)
41. Tversky, A., Kahneman, D.: Judgment Under Uncertainty: Heuristics and Biases. Science 185, 1124-1131 (1974)
42. Jentzsch, N.: Auctioning Privacy-Sensitive Goods. Annual Privacy Forum, pp. 133-142. Springer (2014)
43. Staiano, J., Oliver, N., Lepri, B., de Oliveira, R., Caraviello, M., Sebe, N.: Money Walks: A Human-Centric Study on the Economics of Personal Mobile Data. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 583-594. ACM (2014)
44. Zimmerman, D.W., Zumbo, B.D.: Relative Power of the Wilcoxon Test, the Friedman Test, and Repeated-Measures ANOVA on Ranks. The Journal of Experimental Education 62, 75-86 (1993)
45. Imam, A., Mohammed, U., Moses Abanyam, C.: On Consistency and Limitation of Paired T-Test, Sign and Wilcoxon Sign Rank Test. IOSR Journal of Mathematics 10, 1-6 (2014)
46. Rosenthal, R.: Parametric Measures of Effect Size. In: Cooper, H., Hedges, L.V. (eds.) The Handbook of Research Synthesis, vol. 621, pp. 231-244. New York: Russell Sage Foundation. (1994)
47. Cohen, J.: A Power Primer. Psychological Bulletin 112, 155 (1992)
48. Yim, O., Ramdeen, K.T.: Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. The Quantitative Methods for Psychology 11, 8-21 (2015)
49. Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Hoboken, New Jersey (2009)
50. Foap, https://www.foap.com/terms (Accessed: 16.08.2018)
51. Candidlyimages, https://www.candidlyimages.com/About/FAQ (Accessed: 16.08.2018)
52. DW, https://www.dw.com/en/taxes-coming-to-big-data-in-germany/a-43972540 (Accessed: 24.07.2018)
53. Brustein, J., http://www.nytimes.com/2012/02/13/technology/start-ups-aim-to-help-users-put-a-price-on-their-personal-data.html?_r=1&ref=technology (Accessed: 03.11.2017)
54. Facebook, https://www.facebook.com/terms.php (Accessed: 28.08.2018)
55. Instagram, https://help.instagram.com/581066165581870/ (Accessed: 28.08.2018)

Justification of Mass Surveillance: A Quantitative Study

Jakob Wirth[1], Christian Maier[1] and Sven Laumer[2]

[1] University of Bamberg, Department of Information Systems and Services, Bamberg, Germany
{jakob.wirth,christian.maier}@uni-bamberg.de
[2] Friedrich-Alexander-Universität Erlangen Nürnberg, Schöller Endowed Chair of Information Systems (Digitalization in Business and Society), Nuremberg, Germany
{sven.laumer}@fau.de

Abstract. Online mass surveillance by governmental organizations is omnipresent. Even though this results in a loss of privacy and further negative outcomes for individuals, a majority is justifying mass surveillance. Understanding, why this is the case is among others important for individuals, who want to decrease justification of mass surveillance. Therefore, this study aims to uncover the factors that drive this justification. Drawing on system justification theory we consider mass surveillance to be a political arrangement. Five factors were identified that potentially drive individuals' justification of mass surveillance which are among others, perceived privacy control or perceived security. A quantitative study was carried out and the results support most of our hypotheses. With our results, we contribute to the privacy-related area in the domain of IS, by indicating that individuals consider mass surveillance not necessarily to be bad and by giving advice on how to alter the level of justification.

Keywords: mass surveillance, system justification theory, privacy, nothing to hide, justification

1 Introduction

Governmental organizations have created a system of online mass surveillance. That means, these organizations analyze a vast size of information sent over the Internet without suspicion [1]. The result of this mass surveillance is that a majority of individuals thinks they are having almost no privacy anymore [2]. This reduced privacy results in disadvantages such as profiling, manipulation [3] or chilling effects [4]. For example, among others, individuals do not search online for certain issues such as 'terrorism' anymore because they are afraid becoming a target of governmental organizations [4]. To protect one's own privacy, one might thus assume that individuals reject mass surveillance e.g. by protesting against it.

However, although individuals rate their privacy as very important [2], surveys indicate that more individuals are justifying, i.e. defending and warranting, mass

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

surveillance rather than rejecting it [5, 6], e.g. by protesting against it [7]. What is good for governmental organizations, who have an interest to keep mass surveillance [8], is bad for individuals' privacy [9].

Therefore, both parties – individuals and governmental organizations – need to understand what factors drive individuals' justification of mass surveillance. Then both might have the chance to either change these factors to change the level of justification or to keep these factors to keep the level of justification. Either way, for both parties it is necessary to understand:

What factors drive individuals to justify mass surveillance?

To answer the research question we rely on the one hand on previous research on mass surveillance [6, 10–13]. There, perceived security has been suggested as one of the main factors determining justification of mass surveillance [10, 11, 13]. Yet, this factor has only been suggested and has not been supported scientifically. On the other hand, mass surveillance is a method that is used to monitor a population of individuals, approved and ordered by governmental organizations [14]. Although also companies from the private sector can use mass surveillance technologies, we focus on mass surveillance being a political arrangement as an ordinance by the government. To find out why individuals justify mass surveillance, typical theories from information system (IS) research in the privacy domain cannot be applied [15]. This is because these theories usually focus on an actual IS. In this study, it is more about mass surveillance as a political arrangement. That is why in this research study we refer to system justification theory (SJT) [16–18].

Thus, in the following, we will provide information on SJT as well as mass surveillance and discuss both research streams together to understand mass surveillance justification. A quantitative study is then carried out, using workers from Amazon Mechanical Turk (mTurk). Based on the results, we then contribute to theory by providing factors that are important in a mass surveillance context.

2 Theoretical Background

We next provide information on the system justification theory (SJT) [16, 17]. We continue with research on mass surveillance before carving out the research gap.

2.1 System Justification Theory

System justification: The justification of a system means that individuals defend, bolster or warrant a system [18]. A system is defined as existing social, economic, and political institutions and arrangements. It can range from small-scale systems, such as a nuclear family up to an entire nation [19]. System justification is therefore defined as defending, bolstering or warranting existing social, economic, and political institutions and arrangements.

SJT is a theory to explain why individuals justify a system [17]. Thereby, alternatives of the status quo are degraded. Individuals do so because they have social and psychological needs to consider a system as being legitimate and as being good, fair, natural, desirable, and inevitable [16–18]. They also want to reduce anxiety, guilt, dissonance, discomfort and uncertainty [17].

Prior privacy-research in the domain of IS has not been using SJT in their research studies [15]. In other research domains, SJT has been used to explain justification of different systems including political systems [17], financial work systems [18] or economic systems such as meritocracies [20]. Three factors mainly influence system justification: the perceived powerlessness, the perceived need for order and stability and the perceived dangerousness of the world.

Perceived powerlessness: Power refers to having asymmetric control over particular resources [21]. Individuals, who are in a status of power, are able to process information and to make decision, with the goal to maintain the current position in the system. The more powerful an individual is the more she has access to resources to control her own position [21]. Previous research on SJT has especially researched on individuals who are in a status of perceived powerlessness, i.e. on those individuals who do not have control over particular resources because of living in a particular system [18]. SJT thereby indicates that powerless individuals still justify the system and sometimes even more justify it than powerful individuals, although it is the system, that reduces their power. For example, individuals, who are financially deprived by the current system nevertheless believe in meritocracies [20].

Perceived need for order and stability*:* Individuals, who have a perceived need for order and stability, have the perception that the environment they live in should be orderly, well-structured and unambiguous. If decisions are made they have a desire to stick with these decisions [17].

Perceived dangerousness of the world: Individuals, who think that they live in a dangerous world, have the perception that life is fiercely competitive and ultimately believe that others are constantly threatening to harm them [22]. These individuals have a heightened sensitivity to potential dangers in the social environment, including threats of violence, and terrorism, and have permanent concerns about being in danger [17].

To theorize the relation between system justification theory and mass surveillance, we provide detailed information on mass surveillance in the following section.

2.2 Related Work on Mass Surveillance and Research Gaps

Mass surveillance is generally defined as any method, that collects information of a population, without any attempts to limit the surveillance to a particular individual, but rather to monitor an entire group of individuals [14]. Mass surveillance in an online context has become a major topic in society[1], through revelations proving evidence on how the National Security Agency (NSA) of the USA conduct mass surveillance. Yet, also governmental organizations of other countries conduct mass surveillance on individuals worldwide. It is known that governmental organizations worldwide,

[1] Several EU projects have been conducted to better understand the issue of mass surveillance (please see http://www.projectpact.eu/; http://surprise-project.eu/ and https://cordis.europa.eu/project/rcn/102282_de.html).

essentially store and examine a vast amount of information, which is sent over the Internet, without suspicion [1].

Mass surveillance thereby violates individuals' privacy, whereas privacy is a multi-facet concept [23]. However, in IS research, privacy is usually defined as the amount of control an individual has over her personal information [24], expressed by privacy control [25]. Since mass surveillance takes place by examining information, individuals have often lost control over their information, and thus also over their privacy.

Through mass surveillance, individuals suffer from several disadvantages. For example, individuals search less for delicate issues such as 'terrorists' on the Internet because they are afraid of becoming a target of governmental organizations [4]. Also, discrimination, profiling or manipulation can occur through mass surveillance [3]. Still, it is indicated that more individuals justify mass surveillance rather than oppose it [5], e.g. by protesting against it.

To find out, in how far previous literature has been researching on the topic of mass surveillance justification, we conducted a literature review[2], and identified several articles [6, 10–13, 26–29]. The studies cover topics such as coping with mass surveillance or factors leading to the disapproval of mass surveillance. Still, although covering different issues, all articles commonly suggest two factors that need to be considered when researching on mass surveillance:

Perceived security: Several studies have indicated that increased security in everyday life through mass surveillance might be an important factor why individuals justify mass surveillance. Thereby, it is about the perception of individuals that security is enhanced through mass surveillance [10, 11, 13, 28, 29]. This is in line with governmental organizations which state that through mass surveillance, objective security, such as less terrorist activities, is increased [8].

Nothing to hide: Several studies introduce the concept of 'I have nothing to hide'. This refers to individuals, who state that they have no problem with governmental organizations examining their information, because they think have no information to conceal [10, 11, 27]. Rather they think that only criminals would have something to hide, which would then justify it to spy on it [10].

Research gap: In sum, scholars have already researched on individuals' attitude towards mass surveillance and have therefore strengthened our understanding on this topic. They have particularly indicated that perceived security and the concept of 'nothing to hide' could be important factors in the context of mass surveillance [10, 11, 13, 28, 29]. However, there is scarcity on how these or other factors of the SJT lead to the justification of mass surveillance. We therefore theorize on justification of mass surveillance based on the SJT and the mentioned research on mass surveillance.

3 Theorizing on Justification of Mass Surveillance

In this research study, we use SJT to find out what determines an individual to justify mass surveillance. However, when using a rather general theory in a more particular

[2] The literature review covered the AIS basket of eight, ICIS, HICSS, ECIS and WI proceedings, alongside with a general search on Google Scholar.

context, it is usually inevitable to adapt the theory to the particular context [30]. Therefore, when using SJT to explain mass surveillance, we will on the one hand adapt the existing concepts when necessary. On the other hand, we will add two concepts that are suggested by previous literature on mass surveillance [6, 10–13, 26–29] to better explain the dependent variable.

Thereby, we consider mass surveillance to be a system in the sense of the SJT. That means we consider mass surveillance as an existing political arrangement. Individuals, using the Internet, live in the system of mass surveillance, because through mass surveillance as a political arrangement, they are monitored while using the Internet. Therefore, in light of SJT, justification of a system refers to **justification of mass surveillance**. Based on previous research on SJT and mass surveillance, we define it as the motivation of individuals to defend, bolster or warrant existing political arrangements, allowing mass surveillance, conducted by governmental organizations such as intelligence agencies [16, 29].

Furthermore, we include powerlessness of individuals, which needs to be adapted to the particular context [30]. For example, an individual who is smart but has no physical strength, is powerful on the resource 'information' but powerless on the resource 'strength'. In this study, we research in the context of mass surveillance and privacy of individuals. Privacy of individuals is defined as having control over ones' own personal information, expressed by privacy control [24, 25]. Mass surveillance – the system in this study – is reducing privacy and thus the amount of control over privacy. In SJT, powerlessness refers to the power individuals have but that is reduced by the system. Therefore, in this study, since mass surveillance reduces the power over one's own privacy, we conceptualize powerlessness as the **perceived loss of privacy control.**

We also include both perceptions **perceived need for order and stability** as well as the **perceived dangerousness of the world**. Both perceptions do not need to be adapted to a particular context as they are independent of the situation [17].

Besides the components of the SJT, previous research on mass surveillance suggests two factors, that need to be considered when conducting research on mass surveillance: On the one hand, many individuals think that mass surveillance increases security [28, 29]. This could have an influence on individuals' justification of mass surveillance [13]. We therefore include the component **perceived security** and define it as the perception of an individual that through the conduction of mass surveillance there will be increased security, e.g. less crime or terrorist activities [28, 29]. On the other hand, the argument of many people that they have **nothing to hide** is also suggested to be an important concept in the context of mass surveillance [10, 11]. We therefore include the concept of nothing to hide, and define it as the perception of an individual that she has no information to conceal in front of governmental organizations [10, 11].

Based on these concepts, we build our research model. The research model also includes **prior disclosure** as a control variable. This concept is defined as the amount of information, individuals have disclosed in the past on the Internet [31]. With this variable, we aim to control for, in how the participants have disclosed information on the Internet at all, to account for in how far mass surveillance is even an issue for them. In case they have not disclosed information on the Internet before, it might influence their justification of mass surveillance. Furthermore, as has also been done by previous privacy research [e.g., 31] we also controlled for gender and age. This is because previous research has shown that gender [32] as well as age [24] are variables which often influence attitudes of individuals in a privacy context. Therefore, it is recommended to generally control for them. Hypotheses for the research model are explicated in the following.

Individuals, who are high on perceived loss of privacy control have the perception that they are not able to govern their own personal information. Mass surveillance is a factor, which decreases privacy control [26, 29], because a vast amount of information is captured, stored and examined by governmental organizations, without the control of the individual [1, 29]. Previous research has shown that individuals, who do not have access to needed resources and who are therefore powerless, paradoxically more justify the system, which causes their powerlessness [18]. Individuals, who are high on perceived loss of privacy control, i.e. who are in a less powerful state, will therefore also more justify mass surveillance, which causes their powerlessness. They do so to better maintain a positive image about their situation even if it is not positive at all [18]. We hypothesize:

H1: Perceived loss of privacy control increases justification of mass surveillance.

Having a high need for order and stability means that these individuals want to maintain their status quo [17]. Mass surveillance is considered to be the system, which represents the current status quo. Individuals do not want mass surveillance to stop, because then the status quo is changed, they might need to adapt, and uncertainty arises [33]. This is because one does not know what would happen after mass surveillance has stopped. For example, less security or publishing the results of mass surveillance, would all change the current status quo. Continuing mass surveillance as it is, would at least maintain the status quo even at the expense of decreased privacy. Hence, individuals who have a high need for order and stability, more justify mass surveillance. This is also aligned with the SJT [17] and previous research, suggesting that the need for social order can result in the justification of mass surveillance [28, 29]. We hypothesize:

H2: Perceived need for order and stability increases justification of mass surveillance.

Individuals, who think that they are living in a dangerous world, do see the world as threatening. They think that through certain events, they are under severe danger causing harm. For example, crime activities or terrorist attacks are those events, which increase the perception of severe danger to those individuals [34]. Individuals, who think that they live in a dangerous world, do have that perception across different situations. However, they welcome practices which would reduce their perception, and which fights against that dangerous world. Mass surveillance is said to fight against the dangerous world [26]. This will make them more likely to justify it, which is also aligned with basic premises of the SJT [17]. We hypothesize:

H3: Perceived dangerousness of the world increases justification of mass surveillance.

If individuals think, that there will be more security through mass surveillance, they will also be more likely to justify mass surveillance. This is because increased security is seen as one of the main benefits out of mass surveillance [13]. For example, it is said that through mass surveillance, crime or terrorist activities are reduced, by identifying subjects and forestalling potential crime or terrorist activities [8]. To gain that benefit, individuals will be more likely to justify mass surveillance, which is also suggested by previous research [10, 11, 13, 28, 29]. We hypothesize:

H4: Perceived security increases justification of mass surveillance.

Individuals, who have something to hide, do not want their information to be read by others, e.g., governmental organizations [10, 11]. Therefore, they value their privacy in terms of that they want to control their information. Mass surveillance is against individuals, who have something to hide, because mass surveillance allows governmental organizations to jeopardize individuals' privacy, and to read their private information [10, 11, 27]. Hence, individuals, who have something to hide, will be less likely to justify mass surveillance. The other way around, individuals, who state they have nothing to hide, will therefore be more likely to justify mass surveillance. They see less disadvantages from mass surveillance, and rather think that as long as they do not do something wrong, governmental organizations should be allowed to examine information sent over the Internet [10]. We hypothesize:

H5: Nothing to hide increases justification of mass surveillance.

To evaluate our research model (see Figure 2) we conducted a quantitative study. The methodology therefor is explained in the following section.

4 Methodology

We aim to research on why individuals justify mass surveillance. To do so we conducted a quantitative survey. The used items are depicted in Table 1. To conduct the survey, we used Amazon Mechanical Turk (MTurk) because it is considered to be equivalent to similar data collection methods [35] and it has also been successfully used in a privacy context [36]. We followed the guidelines of previous research to conduct the survey, e.g., by only letting participants take part which have a very high number of already completed tasks. We told workers of MTurk that the maximum time for completing the task will be 12 minutes. Payment was $0.25 for each worker. 141 participants took party in our survey, whereas only these participants who provided answers to at least 90 percent of the questions were kept. All others were removed. After cleaning the data, we ended up with 135 participants in total. Participants were

30.88 years old on average, with a standard deviation of 9.64 years. 38 female participants and 97 male participants took part in our survey.

Table 1. Items and Loadings

Construct	Item	Loadings
Justification of mass surveillance [13]	Intelligence agencies needs to have access to individual bank accounts.	0.840
	Intelligence agencies need wiretapping authority.	0.923
	Intelligence agencies needs to have authority to use high tech surveillance tools for Internet eavesdropping.	0.909
Perceived loss of privacy control [25]	I do not have control over who can get access to my personal information I provided on the Internet.	0.727
	I do not have control over how personal information is used by other parties on the Internet.	0.991
	I do not have control over my personal information provided on the Internet.	0.816
Need for order and stability [37]	I enjoy having a clear and structured mode of life.	0.868
	I like to have a place for everything and everything in its place.	0.871
	I find that a consistent routine enables me to enjoy life more.	0.789
Perceived dangerousness of the world [34]	Every day as society becomes more lawless and bestial, a person's chances of being robbed, assaulted, and even murdered go up and up.	0.907
	It seems that every year there are fewer and fewer truly respectable people, and more and more persons with no morals at all who threaten everyone else.	0.900
	My knowledge and experience tell me that the social world we live in is basically a dangerous and unpredictable place, in which good, decent, and moral people's values and way of life are threatened and disrupted by bad people.	0.844
Perceived security [38, 39]	Mass surveillance by intelligence agencies is good for protecting our country.	0.918
	When conducting mass surveillance, our country is more likely to be protected.	0.951
	Continuous mass surveillance by intelligence agencies lessens the chances of terrorist attacks.	0.902
Nothing to hide [40] Last item was self-developed	People who obey the law have nothing to fear from mass surveillance.	0.945
	Only criminals have any reason to be afraid of mass surveillance.	0.920
	I do not have anything to hide when it comes to mass surveillance by intelligence agencies.	0.678 (dropped)

For the analysis of the survey, we used SmartPLS 3.2.6. The reason is that we also asked for negative perceptions such as loss of privacy control which can skew results [41]. Details and results are given in the following section.

5 Results

To test our research model, we first accounted for common method bias (CMB) followed by the evaluation of the measurement model and the structural model.

Common method bias: We first conducted the Harman's Single-Factor Test, which indicates that 30.46 percent is explained by one factor which shows no indication of CMB [42]. We also accounted for the Unmeasured Latent Marker Construct techniques. The average R^2 with the CMB factor is 0.0027 higher than without the CMB factor. As the R^2 excluding the CMB factor is 0.79135, the ratio is 1:293. Therefore, CMB seems to be no issue [43].

Measurement model: To account for a valid measurement model, we checked on the following criteria: 1) *Indicator reliability*. The threshold of each indicator should be 0.707 to account for at least 50 percent of the variance of the latent variable [44]. As shown in Table 1, this was the case except for one item of 'Nothing to hide' which was dropped in the further evaluation. 2) *Composite reliability (CR)* and *average variance extracted (AVE)*. CR should be above 0.7 and AVE should be above 0.5 [45] which is both the case as one can see in Table 2. 3) *Discriminant validity*. To make sure that all constructs differ from each other, the square root of AVE needs to be greater than the correlation between the constructs [45, 46]. This is also the case as depicted in Table 2. We also computed the heterotrait-monotrait ratio [HTMT, 47]. When using the most conservative approach $HTMT_{0.85,}$ we see no signs of discriminant validity since the highest value is between perceived security and 'Nothing to hide' with 0.742. As all requirements are fulfilled, we conclude that our measurement model is valid.

Table 2. AVE, CR and bivariate correlations

	Mean	St.d.	AVE	CR	1	2	3	4	5	6	7	8	9
1 Justification of mass surveillance	3.48	1.89	0.794	0.920	**0.891**								
2 Perceived loss of privacy control	4.98	1.63	0.725	0.886	-0.138	**0.852**							
3 Need for order and stability	5.07	1.33	0.712	0.881	0.250	-0.004	**0.844**						
4 Perceived dangerousness of the world	4.19	1.53	0.782	0.915	0.478	-0.060	0.289	**0.884**					
5 Perceived security	4.37	1.73	0.854	0.946	0.728	-0.125	0.099	0.475	**0.924**				
6 Nothing to hide	4.22	2.03	0.908	0.952	0.601	-0.089	0.125	0.450	0.742	**0.953**			
7 Prior disclosure	3.96	1.74	0.790	0.919	0.187	0.096	-0.031	0.248	0.290	0.161	**0.889**		
8 Age	30.88	9.64	n/a	n/a	-0.010	0.122	0.007	0.098	0.119	0.163	-0.099	**n/a**	
9 Gender (1=male, 2 = female)	1.28	0.45	n/a	n/a	0.090	-0.025	0.122	-0.003	0.079	0.026	0.109	0.048	**n/a**

The diagonal bold values represent the square root of the AVE of the corresponding constructs.
n/a cannot be computed because these are single-item constructs

Structural model: To evaluate the structural model, we checked on the variance extracted (R^2) of the dependent variable as well as the significance level of each path coefficient. Please see Figure 1 for an overview.

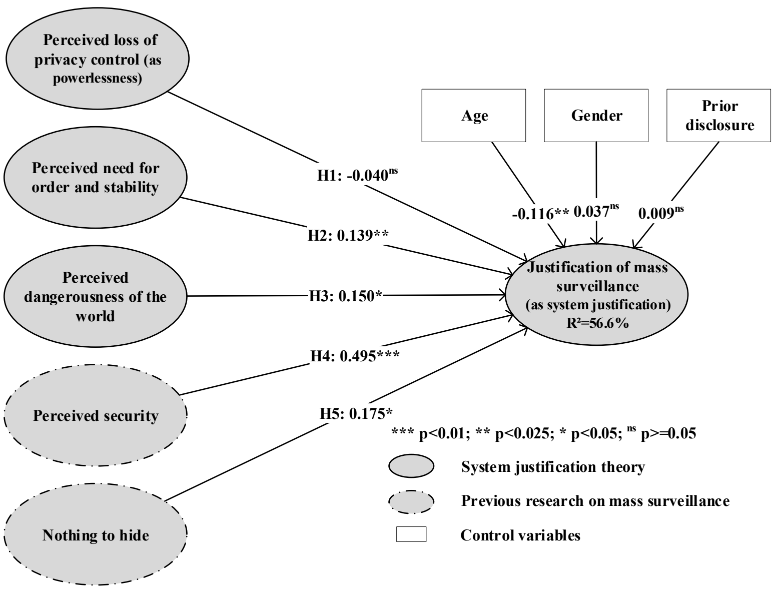


Figure 1. Structural model

The results reveal that 56.6 percent of the variance of justification of mass surveillance is explained. H1 is not supported as the path coefficient is not significant.

Besides, the results support all other remaining hypotheses, i.e. H2, H3, H4 as well as H5. Moreover, the results indicate that gender and prior disclosure as control variables, do not significantly influence justification of mass surveillance. However, age as a control variable has a significant, negative impact. This means that the older the individual, the less likely the justification of mass surveillance.

These results have implications for both, theory and practice, which will be discussed in the following section.

6 Discussion

A majority of individuals is justifying mass surveillance by governmental organizations [5, 6] despite a reduce in one's own privacy [2]. This can lead to further challenges to the individual such as profiling, manipulation [3] or chilling effects [4]. Finding out, why individuals justify mass surveillance, is important for the individuals themselves, in case they want to change their attitude on mass surveillance. Also, for governmental institutions who have an interest in keeping mass surveillance, the results can be important. Scholars have thus already put some research on mass surveillance [10–13]. Yet, research on justification of mass surveillance is rather scarce [15].

In this research study, we therefore concentrated on factors driving justification of mass surveillance. We consider mass surveillance to be a political arrangement, and therefore to be a system in the sense of the system justification theory (SJT) [16]. To use SJT in our research study, we follow previous recommendations to adapt a general theory to a particular context [30], by adapting existing concepts and by adding new concepts from literature on mass surveillance [10, 11, 13, 28, 29]. After having evaluated our research model, we are able to answer our research question which is *what factors drive individuals to justify mass surveillance?*

Our results indicate that perceived security is the major concept increasing justification of mass surveillance. Hence, if individuals have the perception that through mass surveillance, security is about to increase, then they are more likely to justify it. Also, individuals who have nothing to hide have a higher probability to justify mass surveillance. In addition to these two concepts from research on mass surveillance, two factors from SJT are also drivers of justification of mass surveillance: Individuals, who have a higher need for order and stability as well as individuals who perceive the world to be dangerous, are more likely to justify mass surveillance. Besides, perceived loss of privacy control has no influence on justification of mass surveillance. That means, even if individuals have the perception that through mass surveillance, they have lost control over their privacy, it does not have an effect on their justification of mass surveillance. These results have implications for theory in the area of IS, and specifically in the privacy-related domain, as well as for practice.

6.1 Implications for Theory

Mass surveillance is a system in the sense of the SJT: SJT has been used in a variety of settings outside of the IS domain, predominantly in the area of political science [17, 18, 20]. With this research study, we bring SJT to the area of IS research, by conceptualizing mass surveillance to be a political arrangement, and thus a system in the sense of SJT. Scholars, researching on justification of mass surveillance, can therefore also rely on SJT in their future research studies. With this, they are able to rely on additional factors [19] that could be used with SJT, to gain a further understanding of justification of mass surveillance.

Individuals who have a need for order and stability consider mass surveillance to be the current status quo: Individuals who are having a higher need for order and stability want to maintain their current status quo [17]. Since these individuals are more likely to justify mass surveillance, we conclude that they consider mass surveillance to be the current status quo, which they want to maintain. Based on this implication, scholars could use other theories, such as status quo bias theory which has already been applied in IS research [48] or related factors such as right-wing authoritarianism or conservatism [34], to further understand justification of mass surveillance.

Mass surveillance does not have to be a bad thing for individuals: Previous research, including this study, has assumed that mass surveillance is perceived to be negative by individuals, due to privacy issues [e.g. 10]. However, our results indicate that individuals do not necessarily consider mass surveillance to be a bad thing. Individuals think that through mass surveillance, the world might become less dangerous, and perceived security might increase, which is why they justify it. Especially the latter is a factor, that has often been suggested by previous research [13, 26, 28, 29]. However, to the best of our knowledge, we are the first ones to explicitly demonstrate that impact on justification of mass surveillance. Hence, if scholars in the domain of IS, aim to take mass surveillance as something individuals perceive to be negative, or aim to use it to fear individuals [38], they should put caution on it. Researchers should rather check in how far individuals really consider mass surveillance to be a bad thing or if they do not more consider it to be positive.

Loss of privacy control has no impact on justification of mass surveillance: One explanation for that could be, that loss of privacy control does not adequately reflect powerlessness. Powerlessness might also refer to other concepts, e.g., when it comes to the question of who is able to examine information. The majority of individuals is producing information, yet, only a minority of individuals is powerful enough to examine that information. Also, other conceptualizations, such as political power, might better serve as the conceptualization of powerlessness. Another explanation for the non-supported hypothesis would assume, that loss of privacy control does correctly reflect powerlessness. However, powerlessness has no impact on justification in a system of mass surveillance. Previous research on SJT has shown that powerlessness does not always have to lead to more justification of the system [19]. Different reasons have been provided, e.g. when individuals are more ego-centered [18]. Scholars could go deeper into that issue, by applying SJT, and use explanations for the non-significant effects of powerlessness. Independent of that, the results imply for scholars in the domain of IS, that when researching on mass surveillance outside of SJT, a loss of privacy control could be left out in a research model.

Besides theoretical implications, this study also provides practical implications.

6.2 Implications for Practice

Two practical implications arise from this study:

1) Perceived security needs to be aligned with objective security. In case individuals want to change their positive attitude on justification of mass surveillance, they might want to ask themselves, in how far mass surveillance actually increases security. For example, so far, evidence on increased objective security through mass surveillance is rather scarce [49]. Therefore, we recommend individuals to check on, in how far their level of perceived security is aligned with the actual level of objective security through mass surveillance.

2) Individuals who have the need for order and stability consider mass surveillance to be current status quo. Changing the status quo will bring additional problems. Therefore, in case governmental organizations are focusing on individuals, who have a high need for order and stability, they could try to just let things going without changing anything. This is because individuals often try to keep the current status quo [48].

After having discussed the implications of this study, we will now focus on limitations and future research possibilities.

6.3 Limitations and Future Research

This research study uses prior disclosure as a variable to control for, in how far the amount of information disclosed, has an impact on justification of mass surveillance. The results indicate no impact, yet, future research could still concentrate on other demographic variables of the participants, such as the country they are living in, to find out in how far there are differences in the justification of mass surveillance. Furthermore, age as a control variable indicates that the older the individual the less likely she will be to justify mass surveillance. We did not hypothesize on that relationship, yet, future research might more elaborate on these thoughts. One reason could be that older individuals know more how life is without the Internet, and without online mass surveillance, and therefore value that more.

Besides, this research study has concentrated on factors, driving the justification of mass surveillance. If one wants to actually change the level of justification, one could focus on these factors. Future research could thus try to find out, what determines the level of these factors, to then present actual recommendations on how to influence the level of justification of mass surveillance. Furthermore, future research could also expand the definition of mass surveillance, to not only focus on governmental organizations but also include private companies which might also use mass surveillance technologies [e.g., 50]. This could also be done, by focusing on designated journals, which are outside the domain of IS, but which particularly focus on mass surveillance. Future research could on the other hand also focus on the term governmental organizations, by better specifying that term and digging deeper into what particular governmental organizations conduct mass surveillance. In addition, this study has focused on studies in the domain of IS. Future research could also elaborate on other sources outside this domain. Plus, this research study has analyzed the results using a PLS approach. However, other approaches such as the ordinary least squares approach (OLS), might also be suitable. Therefore, future research might also consider using other methodologies than PLS, such as OLS.

In sum, this study has researched on factors, driving justification of mass surveillance. Based on SJT and previous research on mass surveillance, several drivers of mass surveillance justification have been provided, alongside with implications for theory and practice.

References

1. Gidda, M.: Edward Snowden and the NSA files – timeline, http://www.theguardian.com/world/2013/jun/23/edward-snowden-nsa-files-timeline (Accessed: 09.02.2016)
2. Rainie, L.: The state of privacy in post-Snowden America, http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/ (Accessed: 09.03.2017)
3. Acquisti, A., Brandimarte, L., Loewenstein, G.: Privacy and human behavior in the age of information. Science 347, 509–514 (2015)
4. Penney, J.: Chilling Effects: Online Surveillance and Wikipedia Use. Berkeley Technology Law Journal 31, 117–182 (2016)
5. PeWResearchCenter: Few See Adequate Limits on NSA Surveillance Program, http://www.people-press.org/2013/07/26/few-see-adequate-limits-on-nsa-surveillance-program/ (Accessed: 04.05.2017)
6. Reddick, C.G., Chatfield, A.T., Jaramillo, P.A.: Public opinion on National Security Agency surveillance programs. A multi-method approach. Government Information Quarterly 32, 129–141 (2015)
7. Bump, P.: The Majority of Americans Still Don't Care About the NSA Spying on Them, http://www.thewire.com/politics/2013/06/nsa-spying-poll-pew/66093/ (Accessed: 30.08.2016)
8. McAskill, E.: After Charlie Hebdo attack, do spy agencies need more powers?, https://www.theguardian.com/uk-news/2015/jan/09/do-spy-agencies-need-more-surveillance-powers (Accessed: 01.06.2017)
9. Siner, E.: The Internet Flexes Political Muscle With Anti-NSA Protest, https://www.npr.org/sections/alltechconsidered/2014/02/10/274901562/the-internet-flexes-political-muscle-with-anti-nsa-protest?t=1531821946312 (Accessed: 17.07.2018)
10. Solove, D.J.: I've got nothing to hide and other misunderstandings of privacy. San Diego L. Rev. 44 (2007)
11. Solove, D.J.: Why privacy matters even if you have 'nothing to hide'. Chronicle of Higher Education 15 (2011)
12. Moshki, H., Barki, H.: An Exploratory Study on Behavioral and Emotional Coping with IT-Enabled Government Surveillance. In: Proceedings of the 49th Annual Hawaii International Conference on System Sciences. 5-8 January 2016, Kauai, Hawaii, pp. 3636–3645. IEEE, Piscataway, NJ (2016)
13. Dinev, T., Hart, P., Mullen, M.R.: Internet privacy concerns and beliefs about government surveillance – An empirical investigation. The Journal of Strategic Information Systems 17, 214–233 (2008)
14. Privacy International: Mass Surveillance, https://www.privacyinternational.org/node/52 (Accessed: 21.04.2017)

15. Wirth, J.: Dependent Variables in the Privacy-Related Field: A Descriptive Literature Review. In: Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 3658–3667 (2018)
16. Jost, J.T., Banaji, M.R., Nosek, B.A.: A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo. Political Psychology 25, 881–919 (2004)
17. Jost, J.T., Hunyady, O.: Antecedents and Consequences of System-Justifying Ideologies. Current Directions in Psychological Science 14, 260–265 (2005)
18. van der Toorn, J., Feinberg, M., Jost, J.T., Kay, A.C., Tyler, T.R., Willer, R., Wilmuth, C.: A Sense of Powerlessness Fosters System Justification. Implications for the Legitimation of Authority, Hierarchy, and Government. Political Psychology 36, 93–110 (2015)
19. van der Toorn, J., Jost, J.T.: Twenty years of system justification theory: Introduction to the special issue on “Ideology and system justification processes”. Group Processes & Intergroup Relations 17, 413–419 (2014)
20. Jost, J.T., Pelham, B.W., Sheldon, O., Ni Sullivan, B.: Social inequality and the reduction of ideological dissonance on behalf of the system. Evidence of enhanced system justification among the disadvantaged. Eur. J. Soc. Psychol. 33, 13–36 (2003)
21. Keltner, D., Gruenfeld, D.H., Anderson, C.: Power, approach, and inhibition. Psychological Review 110, 265–284 (2003)
22. Cook, C.L., Li, Y.J., Newell, S.M., Cottrell, C.A., Neel, R.: The world is a scary place. Individual differences in belief in a dangerous world predict specific intergroup prejudices. Group Processes & Intergroup Relations, 836–840 (2016)
23. Koops, B.-J., Newell, B.C., Timan, T., Skorvanek, I., Chokrevski, T., Galic, M.: A Typology of Privacy. University of Pennsylvania Journal of International Law 38, 483–575 (2017)
24. Bélanger, F., Crossler, R.E.: Privacy in the digital age: A review of information privacy research in information systems. MIS Quarterly 35, 1017–1042 (2011)
25. Xu, H., Dinev, T., Smith, H.J., Hart, P.: Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View. In: Proceedings of the Twenty Ninth International Conference on Information Systems. Paris, France (2008)
26. Xu, H., Dinev, T.: The security-liberty balance. Individuals' attitudes towards internet government surveillance. Electronic Government 9, 46–63 (2012)
27. Paine, C., Reips, U.-D., Stieger, S., Joinson, A., Buchanan, T.: Internet users’ perceptions of ‘privacy concerns’ and ‘privacy actions’. International Journal of Human-Computer Studies 65, 526–536 (2007)
28. Lyon, D.: The electronic eye. The rise of surveillance society. University of Minnesota Press, Minneapolis (1994)
29. Lyon, D.: Surveillance society. Monitoring everyday life. Open Univ. Press, Buckingham (2008)
30. Hong, W., Chan, F.K.Y., Thong, J.Y.L., Chasalow, L.C., Dhillon, G.: A Framework and Guidelines for Context-Specific Theorizing in Information Systems Research. Information Systems Research 25, 111–136 (2014)
31. Jiang, Z., Heng, C., Choi, B.C.F.: Privacy Concerns and Privacy-Protective Behavior in Synchronous Online Social Interactions. Information Systems Research 24, 579–595 (2013)
32. Sun, Y., Wang, N., Shen, X.-L., Zhang, J.X.: Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. Computers in Human Behavior 52, 278–292 (2015)
33. Laumer, S., Maier, C., Eckhardt, A., Weitzel, T.: Work routines as an object of resistance during information systems implementations. Theoretical foundation and empirical evidence. Eur J Inf Syst 25, 317–343 (2016)

34. Altemeyer, B.: Right-wing authoritarianism. University of Manitoba Press, Winnipeg (1981)
35. Lowry, P.B., D'Arcy, J., Hammer, B., Moody, G.D.: "Cargo Cult" science in traditional organization and information systems survey research. A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. JSIS 25, 232–240 (2016)
36. Wirth, J., Maier, C., Laumer, S.: The influence of resignation on the privacy calculus in the context of social networking sites: an empirical analysis. In: Proceedings of the Twenty-Sixth European Conference on Information Systems. Portsmouth, UK (2018)
37. Neuberg, S.L., Newsom, J.T.: Personal need for structure. Individual differences in the desire for simpler structure. Journal of Personality and Social Psychology 65, 113–131 (1993)
38. Boss, S.R., Galletta, D.F., Lowry, P.B., Moody, G.D., Polak, P.: What Do Systems Users Have to Fear? Using Fear Appeals to Engender Threats and Fear that Motivate Protective Security Behaviors. MIS Quarterly 39, 837–864 (2015)
39. Johnston, A.C., Warkentin, M.: Fear appeals and information security behaviors: An empirical study. MIS Quarterly 34, 549-A4 (2010)
40. Honess, T., Charman, E.: Closed circuit television in public places. Its acceptability and perceived effectiveness. Home Office Police Dept, London (1992)
41. Turel, O., Serenko, A., Giles, P.: Integrating technology addiction and use: An empirical investigation of online auction users. MIS Quarterly 35, 1043–1062 (2011)
42. Schwarz, A., Rizzuto, T., Carraher-Wolverton, C., Roldan, J.L., Barrera-Barrera, R.: Examining the Impact and Detection of the "Urban Legend" of Common Method Bias. ACM Sigmis Database 48, 93–119 (2017)
43. Liang, H., Saraf, N., Hu, Q., Xue, Y.: Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. MIS Quarterly 31, 59–87 (2007)
44. Carmines, E.G., Zeller, R.A.: Reliability and validity assessment. Sage Publications, Newbury Park, California (2008)
45. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 18, 39–50 (1981)
46. Campbell, D.T., Fiske, D.W.: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56, 81–105 (1959)
47. Henseler, J., Ringle, C.M., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. JAMS 43, 1–21 (2014)
48. Polites, G.L., Karahanna, E.: Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. MIS Quarterly 36, 21–42 (2012)
49. McLaughlin, J.: U.S. Mass Surveillance Has No Record of Thwarting Large Terror Attacks, Regardless of Snowden Leaks, https://theintercept.com/2015/11/17/u-s-mass-surveillance-has-no-record-of-thwarting-large-terror-attacks-regardless-of-snowden-leaks/ (Accessed: 04.05.2017)
50. Zureik, E.: Surveillance, privacy, and the globalization of personal information. International comparisons. McGill-Queen's University Press, Montréal, Ithaca [N.Y.] (2010)

An Exploratory Study of Risk Perception for Data Disclosure to a Network of Firms

Tobias Steudner[1], Thomas Widjaja[1] and Jan H. Schumann[2]

[1] University of Passau, Chair of Business Information Systems, Passau, Germany
{tobias.steudner,thomas.widjaja}@uni-passau.de
[2] University of Passau, Chair of Marketing and Innovation, Passau, Germany
jan.schumann@uni-passau.de

Abstract. Research on the Privacy Calculus, which explains individuals' intention to disclose personal data, mostly focuses on dyadic disclosures in which individuals disclose data to a single firm. So far, little attention has been paid to understand the characteristics of data disclosures to a network of firms. We refer to data sharing of firms in a network as "Business Network Data Exchange" (BNDE). We explore risk perception for data disclosures in a BNDE context based on an exploratory survey. Our results indicate that risk perception for data disclosures in the BNDE context deviates from rational risk perception theory. In particular, individuals perceive the risk to disclose data to a network of two firms as lower than the maximum risk of the separate dyadic data disclosures. These results portend the need for an adapted and nuanced view on perceived risks in this context and have important practical implications for data-sharing among firms.

Keywords: Privacy Calculus, Risk Perception, Business Network Data Exchange

1 Introduction

Most research in the privacy context focuses on situations in which individuals disclose data only to a single firm (e.g., [1]). However, recently more and more firms started to depart from this dyadic consumer-firm relationship and began to share consumer data within a network of firms [2, 3]. In accordance with Bidler et al., we will refer to such procedures as Business Network Data Exchange (BNDE) [4]. An example for BNDE is the music streaming service Spotify: Consumers' data is shared among a network of artists, record labels, and further third parties [5]. Potential differences between dyadic data disclosures and data disclosures in the BNDE context have rarely been examined in the literature. One difference could be the complexity or non-transparency of BNDE situations, which could promote irrational behavior [4, 6–8]. However, the Privacy Calculus, as the dominant theory to explain individuals' intention to disclose personal data by weighing their benefits against their risks, assumes rational behavior [3, 9–11]. As perceived privacy risks are the main factor reducing individuals' intention to disclose in the Privacy Calculus [11–13], we will focus on perceived risks as a key factor in this first approach. We argue that in BNDE data disclosures individuals' risk perception

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

differ from dyadic data disclosures. As a starting point for future research in the BNDE context, we first focus on the question: *Is individuals' risk perception of a data disclosure to a BNDE firm network consisting of two firms higher or lower than the dyadic data disclosure to a single firm?*

2 Theoretical Background on Risks

Privacy risk comprises "the degree to which an individual believes that a high potential for loss is associated with the release of personal information to a firm" [3]. In traditional risk perception theory a rational evaluation is expected as the perceived risk of an outcome is defined as the probability of a certain unfavorable outcome multiplied by the severity of the respective outcome [14–16]. If a risk is constituted by different risk components they are often assumed to be additive [17]. The assumption of rational evaluation as well as the additivity of risks was challenged by the Prospect Theory [18, 19]. According to the Prospect Theory individuals transform benefits and risks into a simpler mental representation and use a function to assign a subjective value (perceived benefits and perceived risk) to them. For this value function, monotony as well as a diminishing marginal value is assumed [18, 19]. Since the main emphasis of our paper is on risks, this means particularly that each added potential loss should increase the perceived risk (monotony), but this effect diminishes the more potential losses already exist (diminishing marginal value). In this short paper we focus on the simplest possible BNDE network consisting of just two firms (see Table 1). According to traditional risk theory as well as Prospect Theory, we expect the risk of a data disclosure to a BNDE network with two firms should be perceived higher than the risk of each of the dyadic data disclosures due to the monotony assumption. The risk increase should be caused by the additional (second) firm which obtains the individuals' data. This leads to our proposition 1a.

On the other hand, mutual control in cooperations plays a major role for firms [20, 21], thus it is possible that individuals' perceived or assumed control distribution among the firms in the network could alter individuals' risk perception as well. Furthermore, cooperations among firms offer several benefits for consumers and firms. For instance, firms can complement each other [22, 23], e.g., with technical security know-how. Additionally, firms can improve their own internal processes due to knowledge transfer in a cooperation [22, 24–26], which could reduce individuals' perceived risk for BNDE data disclosures, e.g., through better standards of conduct. Both was also described by some participants in pre-test interviews. This leads to our proposition 1b.

Proposition 1a / 1b: *The perceived risk for the BNDE data disclosure is perceived higher (Proposition 1a) / lower (Proposition 1b) than the maximum perceived risk of the two respective dyadic data disclosures.*

3 Exploratory Survey

The survey data (61 subjects, 33 male) was collected from end of November 2017 until mid of December 2017 in cooperation with a panel provider. The subjects were over 18 years, live in Germany, and the age distribution of them reflects the age distribution of German Internet users. The within-subject designed survey contained two groups with three scenarios each and a fixed scenario order (see Table 1).

Table 1. Empirical Setup

	Scenario 1 (BNDE)	Scenario 2 (Dyadic)	Scenario 3 (Dyadic)
Group 1 (n=32)	→ High-Risk-Firm → Low-Risk-Firm	→ High-Risk-Firm	→ Low-Risk-Firm
Group 2 (n=29)	→ Low-Risk-Firm → High-Risk-Firm	→ High-Risk-Firm	→ Low-Risk-Firm

The first scenario was a BNDE data disclosure scenario with two existent firms: The subjects saw a screenshot of the respective firm's website and were asked to disclose some personal data (name, address, net-household income, expenses per week, supermarkets they regularly visit, number of persons in the household, phone number). In group 1 (n=32) the subjects disclosed their data to the High-Risk-Firm, which shares the exact same data subsequently not anonymized with the Low-Risk-Firm[1]. This data sharing procedure between the two firms was described in the scenario. In group 2 (n=29) the data was disclosed to the Low-Risk-Firm and then shared with the High-Risk-Firm. In both groups, both firms obtain the same data. The benefit in all three scenarios was identical (20 Euro coupon which could be redeemed in both firms). Two groups were used to control for firm order effects.

After reading the respective scenarios the subjects had to assess their perceived privacy risk (PR) on a seven-point Likert Scale with one being "strongly disagree" and 7 being "strongly agree" based on a scale of Dinev et al. [27] (PR1: It is very risky in this situation to disclose personal information. PR2: There would be high potential for privacy loss associated with data disclosure in this situation. PR3: Personal information could be inappropriately used in this situation. PR4: Providing personal information in this situation would involve many unexpected problems; The average of the items was used as perceived risk). For the High-Risk-Firm a well-known search engine firm and for the Low-Risk-Firm a well-known grocery store were selected. Existing firms were used to make the whole scenario easier to imagine and comprehend. After assessing the risk for the BNDE disclosure (scenario 1) the subjects had to imagine and assess their privacy risk for the two respective dyadic data disclosures without any data sharing to further firms (i.e., scenario 2: disclosure only to the High-Risk-Firm and scenario 3: disclosure only to the Low-Risk-Firm). These two dyadic scenarios were the same in both groups.

4 Results, Discussion, and Outlook

As expected, the risk for the dyadic scenario 2 ($PR_{High\text{-}Risk\text{-}Firm}$) is perceived in both groups significantly higher than the risk for the dyadic scenario 3 ($PR_{Low\text{-}Risk\text{-}Firm}$), thus $PR_{High\text{-}Risk\text{-}Firm}$ is the maximum perceived risk of the two dyadic data disclosures. We used "R" with "EnvStats" for a one-sided paired randomization test for location [28–32] with 100k iterations to test each group separately as well as both groups together (see Table 2, Δ means test statistic). $PR_{BNDE\text{-}DD}$, the perceived risk for scenario 1, was significantly smaller ($\Delta = 40$, $p = 0.001$) than the maximum perceived risk out of the dyadic data disclosures ($PR_{High\text{-}Risk\text{-}Firm}$), thus we reject proposition 1a and accept 1b.

1 We conducted interviews as a pre-test (14 subjects, 11 male, age between 21 and 71) to ensure that the risk level is perceived as different between the two dyadic data disclosures.

Table 2. Preliminary Results

Group \ Scenario	Mean Perceived Risk (Standard Deviation)			One-Sided Paired Randomization Test for	
	PR_{BNDE} Scenario 1	$PR_{High\text{-}Risk\text{-}Firm}$ Scenario 2	$PR_{Low\text{-}Risk\text{-}Firm}$ Scenario 3	$PR_{Low\text{-}Risk\text{-}Firm} < PR_{High\text{-}Risk\text{-}Firm}$	$PR_{BNDE} < PR_{High\text{-}Risk\text{-}Firm}$
Group 1 (n=32)	4.25 (1.52)	5.11 (1.33)	3.52 (1.44)	$\Delta = 50.75$, $p = 0.000$	$\Delta = 27.50$, $p = 0.003$
Group 2 (n=29)	4.90 (1.52)	5.33 (1.33)	4.35 (1.44)	$\Delta = 28.25$, $p = 0.003$	$\Delta = 12.50$, $p = 0.074$
Both groups (n=61)	4.56 (1.61)	5.21 (1.52)	3.92 (1.63)	$\Delta = 79.00$, $p = 0.000$	$\boxed{\Delta = 40.00, p = 0.001}$

Contrary to the monotony assumption of the Prospect Theory, the perceived risk for data disclosures in a BNDE context with two firms is perceived as less risky than the maximum of the two respective dyadic data disclosures. These preliminary results indicate that we observe non-monotony behavior that could be explained by positive cooperation effects which reduce individuals' perceived risk due to individuals' assumption of positive changes in firms' data handling process. Future research could review the monotony assumption of the Prospect Theory in the BNDE context and extend the Privacy Calculus by considering positive cooperation effects. Future research should also investigate whether the changes in individuals' risk perception are an instance of irrational behavior in the context of the Privacy Calculus as suggested by Dinev et al. [9].

Additionally, our preliminary results have important practical implications: When the risk for a BNDE data disclosure is perceived as less risky than one of the separate dyadic disclosures, joint data collections are more effective in total and specifically for firms associated with high risk. Since the perceived risk for BNDE data disclosures is nevertheless expected to be the same or higher than the minimum of the separate dyadic data disclosures, it might be beneficial to investigate redistribution mechanisms among firms to balance their cooperation benefits.

The presented preliminary results should be viewed in the light of their limitations: The sample size is small and the within-design without scenario order randomization could distort the results. Also, our results could hold only for BNDE networks consisting of just two firms. To rule out these possible error sources a new survey with bigger sample size, hypothetical firms, varying network size, and a between-subject design is in work.

In sum, we showed that data disclosures in a BNDE context are an interesting and unexplored field that requires further research. Thus, we aim for a deeper understanding in which BNDE constellations (e.g., firm combinations, network characteristics, etc.,) individuals assume positive changes in the data handling process of the network firms and perceive less risk for the BNDE data disclosure. For this, our exploratory study with the focus on risk perception shall serve as a starting point.

References

1. Li, H., Sarathy, R., Xu, H.: The role of affect and cognition on online consumers' decision to disclose personal information to unfamiliar online vendors. Decis. Support Syst. 51, 434–445 (2011).
2. Madsbjerg, S.: It's Time to Tax Companies for Using Our Personal Data, https://www.nytimes.com/2017/11/14/business/dealbook/taxing-companies-for-using-our-personal-data.html (Accessed: 24.04.2018).
3. Smith, H.J., Dinev, T., Xu, H.: Information privacy research: an interdisciplinary review. MIS Q. 35, 989–1016 (2011).

4. M. Bidler, J. H. Schumann, T. Widjaja: Challenging the Cognitive Privacy Calculus: Affective Reactions in Consumers' Privacy Related Decision Making. Presented at the EMAC Conference. Glasgow, United Kingdom (05. - 01.06.2018).
5. Spotify: Spotify Privacy Policy, https://www.spotify.com/us/legal/privacy-policy/ (Accessed: 23.04.2018).
6. Acquisti, A., Brandimarte, L., Loewenstein, G.: Privacy and human behavior in the age of information. Science. 347, 509–514 (2015).
7. Amos Tversky, Daniel Kahneman: Rational Choice and the Framing of Decisions. J. Bus. 59, 251–278 (1986).
8. Finucane, M.L., Alhakami, A., Slovic, P., Johnson, S.M.: The Affect Heuristic in Judgments of Risks and Benefits. J. Behav. Decis. Mak. 13, 17 (2000).
9. Dinev, T., McConnell, A.R., Smith, H.J.: Research Commentary—Informing Privacy Research Through Information Systems, Psychology, and Behavioral Economics: Thinking Outside the "APCO" Box. Inf. Syst. Res. 26, 639–655 (2015).
10. Laufer, R.S., Wolfe, M.: Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. J. Soc. Issues. 33, 22–42 (1977).
11. Li, Y.: Theories in online information privacy research: A critical review and an integrated framework. Decis. Support Syst. 54, 471–481 (2012).
12. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. Inf. Syst. Res. 15, 336–355 (2004).
13. Xu, H., Teo, H.-H., Tan, B.C.Y.: Predicting the adoption of location-based services: The role of trust and perceived privacy risk. In: ICIS 2005 proceedings. pp. 897–910 (2005).
14. Bauer, R.A.: Consumer Behavior as Risk Taking. In: Risk Taking and Information Handling in Consumer Behavior. pp. 389–398. , Boston: Harvard University (1967).
15. Cunningham, S.: The Major Dimensions of Perceived Risk. In: Risk Taking and Information Handling in Consumer Behavior. pp. 82–108. , Boston: Harvard University (1967).
16. Sieber, J.E., Lanzetta, J.T.: Conflict and conceptual structure as determinants of decision-making behavior1. J. Pers. 32, 622–641 (2006).
17. Peter, J.P., Tarpey, L.X.: A Comparative Analysis of Three Consumer Decision Strategies. J. Consum. Res. 2, 29–37 (1975).
18. Tversky, A., Kahneman, D.: Advances in prospect theory: Cumulative representation of uncertainty. J. Risk Uncertain. 5, 297–323 (1992).
19. Kahneman, D., Tversky, A.: Prospect Theory: An Analysis of Decision under Risk. Econometrica. 47, 263–291 (1979).
20. Killing, J.P.: How to Make a Global Joint Venture Work, https://hbr.org/1982/05/how-to-make-a-global-joint-venture-work (Accessed: 24.04.2018).
21. Ahern, R.: The Role of Strategic Alliances in the International Organization of Industry. Environ. Plan. A. 25, 1229–1246 (1993).
22. Lei, D., Slocum Jr., J.W.: Global Strategy, Competence-Building and Strategic Alliances. Calif. Manage. Rev. 35, 81–97 (1992).
23. Mason, J.C.: Strategic alliances: Partnering for Success. Manage. Rev. 82, 10 (1993).
24. Hamel, G., Doz, Y.L., Prahalad, C.K.: Collaborate with Your Competitors and Win. 8 (1989).
25. Doz, Y.L.: Technology Partnerships between Larger and Smaller Firms: Some Critical Issues. Int. Stud. Manag. Organ. 17, 31–57 (1987).
26. Prahalad, C.K., Doz, Y.L.: The Multinational Mission: Balancing Local Demands and Global Vision. Simon and Schuster (1999).

27. Dinev, T., Xu, H., Smith, J.H., Hart, P.: Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. Eur. J. Inf. Syst. 22, 295–316 (2013).
28. Fisher, R.A.: The Design of Experiments. Des. Exp. (1935).
29. Smucker, M.D., Allan, J., Carterette, B.: A comparison of statistical significance tests for information retrieval evaluation. In: Proceedings of the sixteenth ACM Conference on Information and Knowledge Management. p. 623. ACM Press, Lisbon, Portugal (2007).
30. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. CRC Press (1994).
31. Cohen, P.: Empirical Methods for Artificial Intelligence. 128 (2006).
32. Millard, S.P.: EnvStats: An R package for environmental statistics. Springer, New York (2013).

Track 12:

Umweltinformatik und nachhaltiges Wirtschaften

Track Chairs:

Prof. Dr. Jorge Marx Gómez
Universität Oldenburg

Dr. Alexander Boden
Fraunhofer Institute for Applied Information Technology FIT

Kommunikationsfäden im Nadelöhr – Fachliche Prozessmodellierung der Nachhaltigkeitskommunikation am Kapitalmarkt

Raphaela Helbig[1,2] und Jorge Marx Gómez[2]

[1] Volkswagen Aktiengesellschaft, Wolfsburg, Deutschland
[2] Carl von Ossietzky Universität Oldenburg, Department für Informatik, Abt. Wirtschaftsinformatik/ VLBA, Oldenburg, Deutschland
{raphaela.helbig,jorge.marx.gomez}@uol.de

Abstract. Am Kapitalmarkt steigt der Bedarf einer Stakeholder-orientierten Nachhaltigkeitskommunikation von Unternehmen. Informationssysteme können Unternehmen in einer Stakeholder-orientierten und effizienten Nachhaltigkeitskommunikation unterstützen. Mithilfe einer fachlich deskriptiven Prozessmodellierung in BPMN 2.0 wird der Status quo der informationstechnisch verstandenen Nachhaltigkeitskommunikation beschrieben. Dies erfolgt auf Basis von Fachliteratur durch Abstraktion und Partitionierung. Es wird deutlich, dass Investoren vorrangig auf Primär- und Sekundärinformationen von Intermediären zurückgreifen und weniger auf Primärinformationen direkt von Unternehmen. Das entwickelte Ist-Modell eignet sich als Forschungsgrundlage in der Wirtschaftsinformatik, um beispielsweise die Nachhaltigkeitskommunikation mit einer Plattform zielgruppengerecht zu fördern. Es eröffnet sich ein grundlegendes Forschungs- und Anwendungsgebiet am Kapitalmarkt.

Keywords: Nachhaltigkeitsberichterstattung, SRI, nachhaltige Kapitalanlage, ESG-Research, Nachhaltigkeitsrating

1 Einleitung

Der Markt für nachhaltige Kapitalanlagen wächst und umfasst aktuell weltweit über ein Viertel der professionell verwalteten Kapitalanlagen [1]. Häufig wird die Finanzwirtschaft als Katalysator für die weltpolitisch beabsichtigte nachhaltige Entwicklung bezeichnet. Gemeint ist, dass die Finanzwirtschaft eine ökologisch und ethisch verantwortliche Transformation von Unternehmen beschleunigt [2]. Kapitalanlagen spielen in diesem Zusammenhang eine wichtige Rolle. Am Markt für nachhaltige Kapitalanlagen ist die Nachhaltigkeitskommunikation wesentlich, weil dort Investoren Nachhaltigkeit in ihrer Risikoabschätzungen berücksichtigen. Weiterhin ist die klassische Kapitalmarktkommunikation von Unternehmen mit wachsender Heterogenität der Investoren und mit der Digitalisierung in der Finanzwirtschaft konfrontiert [3]. Es stellt sich die Frage, ob die Nachhaltigkeitskommunikation von Unternehmen mit Blick auf den Kapitalmarkt verbessert werden kann, bspw. unter Einbezug von IT-Systemen.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Der vorliegende Beitrag beschreibt das aktuelle ökonomische Phänomen der Nachhaltigkeitskommunikation börsennotierter Großunternehmen am Kapitalmarkt. Ausgangspunkt ist der theoretische Kommunikationsbegriff aus Nachrichten-Perspektive. Definiert wird Nachhaltigkeitskommunikation somit als Austausch von Nachhaltigkeitsinformationen am Kapitalmarkt. Deutlich wird dabei, dass die Fäden der Nachhaltigkeitskommunikation zwischen Unternehmen und Investoren aktuell vorrangig über Intermediäre laufen. Letztere werden im Beitrag als Nadelöhr bezeichnet.

2 Perspektiven auf den Kommunikationsbegriff

Grundlage der Untersuchung bildet der theoretische Kommunikationsbegriff aus Nachrichten-Perspektive. Zur Diskussion der Ergebnisse wird punktuell der Kommunikationsbegriff aus der Perspektive der Sprechakt-Theorie hinzugezogen.

Die Nachrichten-Perspektive stammt aus der Informatik und betrachtet die Daten- und Informationsübertragung zwischen Sender und Empfänger gemäß dem Kanalmodell von Shannon und Weaver. Demnach entsteht zielführende Kommunikation dann, wenn Nachrichten gleichartig kodiert und dekodiert sowie störungsfrei übertragen werden [4, 5]. Schnittstellen mit der realen Welt sind hierbei Dateneingaben, -zugänge und Berichte [6]. Die Nachrichten-Perspektive versteht Kommunikation lediglich als das Übermitteln von Informationen und Daten. Informationen haben stets einen Zweckbezug und können sowohl maschinell als auch nicht-maschinell verarbeitet sein; Daten hingegen sind maschinell verarbeitet, ein Zweckbezug ist nicht notwendig [5].

Die Perspektive der Sprechakt-Theorie (Sprechakt-Perspektive), die der komplexen Lebenswelt zugewandt ist, definiert Kommunikation als Interaktion, die auf einen Konsens zu Geltungsansprüchen basiert. In den Sozialwissenschaften meint Kommunikation vor allem einen zwischenmenschlichen Prozess. Kommunikation kann sich demnach entwickeln aus einem „wechselseitig aufeinander bezogen (interaktiven) und absichtsvollen (doppelte Intention) kommunikativen Handeln“ [4] und basiert auf Sprachen mit Syntax und Semantik. Die Sprechakt-Theorie nach Austin und Searle, sowie die daran anknüpfende Theorie des Kommunikativen Handelns nach Habermas gehen davon aus, dass sprachliche Äußerungen Handlung sind und Sprache folglich erst Interaktion ermöglicht. Sprechhandlungen setzen jedoch voraus, dass die Inhalte verstanden werden. Die Theorie des Kommunikativen Handelns geht von universalen Bedingungen der Verständigung aus: Verständlichkeit, Wahrheit, Wahrhaftigkeit und Richtigkeit. Diese Geltungsansprüche bieten Ansatz zur Analyse von Verständigungsproblemen [7]. Unter Realbedingungen werden Beteiligte diese Geltungsansprüche nicht jederzeit vollständig anerkennen. Vielmehr werden diese durch soziale Diskurse thematisiert und fortwährend verändert, um schließlich das kommunikative Handeln weiterzuführen [7]. An dieses Kommunikationsverständnis der Sprechakt-Theorie knüpfen Winograd [8] und Flores et al. [6] für ihr auf Interaktion basierendes Systemdesign an. Damit grenzen sie sich von der Nachrichten-Perspektive der Datenübertragung ab, die dem herkömmlichen Entwicklungsansatz technischer Systeme zugrunde liegt [6].

3 Grundlagen und Forschungsstand

3.1 Investoren und nachhaltige Kapitalanlagen

Investoren zeigen zunehmend Interesse an unternehmensrelevanten Nachhaltigkeitsinformationen und nachhaltigen Kapitalanlagen. In den letzten Jahrzehnten ist der Nischenmarkt für Kapitalanlagen gewachsen, die nach ethischen und ökologischen Kriterien investiert werden [9]. Zunehmend nutzen auch Mainstream-Investoren Nachhaltigkeitsdatenbanken [10], bekennen sich mit der Initiative PRI (Principles for Responsible Investment) zu Grundsätzen des verantwortlichen Investierens [11] und unterstützen öffentlich die Abfrage umweltrelevanter Daten bei Unternehmen durch das CDP (Carbon Disclosure Project) [12]. Am Kapitalmarkt gewinnt deshalb bei Investitionsentscheidungen neben den klassisch wirtschaftlichen Anlagezielen (Rentabilität, Sicherheit, Liquidität) ein viertes Anlageziel an Bedeutung – die Nachhaltigkeit [13]. Der Markt für nachhaltige Kapitalanlagen – auch als Markt für Socially Responsible Investment (SRI) bekannt – hat sich aus der Nische heraus zusätzlich, mit unterschiedlichen Ausprägungen, in den Mainstream entwickelt.

Mit dem Anlageziel der Nachhaltigkeit kann eine nachhaltige Entwicklung von Unternehmen gezielt unterstützt werden [2]. Dieser Logik folgend haben Investoren den Anreiz, neben ökonomischer Rendite auch einen immateriellen Mehrwert durch ihre Kapitalanlage zu erzielen. Gleichzeitig setzt dieses Anlageverhalten bei Unternehmen einen Anreiz zur Umsetzung einer nachhaltigen Wertschöpfung, um ihren Marktwert zu maximieren [14]. In einer aktuellen Metastudie von Friede et al. [15] mit über 2000 Einzelstudien wurde ein positiver Zusammenhang von Nachhaltigkeit und finanzieller Performance bestätigt. Zusätzlich zu dieser extrinsischen ökonomischen Motivation ist es auch denkbar, dass Unternehmen intrinsisch motiviert eine nachhaltige Wertschöpfung umsetzen.

Investoren sind heterogen, mit verschieden ausgeprägtem Informationsbedürfnis und unterschiedlichen Investitionsentscheidungen. Es gibt private und institutionelle Investoren. Zu letzteren gehören Pensionsfonds, Versicherungen, Beteiligungs- und Investmentgesellschaften. Für Entscheidungen benötigen Investoren differenzierte Informationen aufgrund unterschiedlicher Ausrichtungen und Zwecke. [3, 16]

Nachhaltigkeit wird am Kapitalmarkt i. d. R. mit dem Akronym ESG zusammengefasst [17]. Diese Abkürzung steht für die Kategorien Ökologie, Soziales und Unternehmensführung. Nach diesen Kategorien bewerten Investoren das Risiko von Kapitalanlagen. Über eine genaue Definition von ESG gibt es jedoch keinen Konsens, da die inhaltliche Ausprägung je nach Nachhaltigkeitsverständnis verschieden ausfällt.

Mit Blick auf das Nachhaltigkeitsverständnis bilden Autoren und Institutionen häufig Abstufungen [9, 18-21]. Der vorliegende Beitrag orientiert sich an dem weitgefassten inklusiven Verständnis der Global Sustainable Investment Alliance [1] und bezeichnet das gesamte Spektrum vereinfachend als nachhaltige Kapitalanlagen. Wird diese erweiterte Definition für nachhaltige Kapitalanlagen verwendet, so sind derzeit weltweit 26% der gesamten verwalteten Kapitalanlagen nachhaltig investiert.

3.2 Praxis der Nachhaltigkeitsberichterstattung und -kommunikation von Unternehmen und Intermediären

Nachhaltigkeitskommunikation im Unternehmenskontext wird in der Praxis aus der Nachrichten-Perspektive gesehen und auch als Nachhaltigkeitsberichterstattung oder -reporting bezeichnet. Gemeint ist der Austausch von nachhaltigkeitsrelevanten Unternehmensinformationen. Seit einiger Zeit stellen Unternehmen Nachhaltigkeitsinformationen für ihre Stakeholder (Anspruchsgruppen) bereit, um ihre Nachhaltigkeitsperformance offenzulegen. Im Zusammenhang mit der Nachhaltigkeit von Unternehmen fällt häufig der Begriff Corporate Social Responsibility (CR, CSR), der das nachhaltige Verhalten und die Verantwortung von Unternehmen umfasst. In der Praxis der Unternehmenskommunikation ist der Nachhaltigkeitsbericht das Hauptmedium und es überwiegt die einseitige Nachhaltigkeitsberichterstattung von Unternehmen mit Multi-Stakeholderansatz. Für die Nachhaltigkeitsberichterstattung gibt es eine Vielzahl an Rahmenwerken und Standards [22], bspw. Global Reporting Initiative (GRI), Global Compact, AA1000 und ISO14001. Nur wenige dieser Rahmenwerke sind speziell auf die Stakeholdergruppe der Investoren ausgerichtet [18, 23-25]. Die weltweit am meisten genutzten Rahmenwerke sind die der GRI [26]. Zunehmend wichtig wird die Orientierung auf Stakeholder in der unternehmerischen Nachhaltigkeitskommunikation, da Stakeholder dieser mit unterschiedlichen Erwartungen entgegentreten [27]. Daher führen Unternehmen Stakeholder-Dialoge zu Nachhaltigkeitsthemen durch, um ihre Berichterstattung mit den Sichtweisen der Stakeholder abzugleichen [28].

Es gibt mehrere Gründe, die Unternehmen zur Kommunikation von unternehmensrelevanten Nachhaltigkeitsinformationen veranlassen. Die Publikation nicht-finanzieller Informationen ist für Großunternehmen durch das deutsche CSR-Richtlinie-Umsetzungsgesetz (CSR-RUG) seit dem Berichtsjahr 2017 zur Pflicht geworden (in Kraft seit 18.4.2017, setzt EU-Richtlinie um). Aus Investorensicht sind besonders die Nachhaltigkeitsinformationen von börsennotierten Großunternehmen interessant, deren Anteile in Streubesitz sind [16, 18]. Weitere Motive von Unternehmen sind nicht-materiell, wie bspw. Glaubwürdigkeit, Vertrauen und Image [29].

In der Nachhaltigkeitskommunikation mit der Stakeholdergruppe der Investoren fungieren häufig Intermediäre als Brücke, um zuverlässige unternehmensrelevante Nachhaltigkeitsinformationen an die Investoren heranzutragen. Zu diesem Zweck erheben Intermediäre – meist unaufgefordert oder seltener mit Beauftragung [18-19] – Informationen über Unternehmen. Zudem verfügen sie über Fachwissen und Erfahrung zur Bewertung der Nachhaltigkeitsperformance. Diese Definition der Intermediäre schließt alle heterogenen Eigenschaften ein, bspw. Unterschiede in Nachhaltigkeitsverständnis, Relevanz, Informationsgrundlage und Produktangebot. Viele Unternehmen teilen die Auffassung, dass ein positives Abschneiden in Nachhaltigkeits-Ratings die immateriellen Vermögenswerte positiv beeinflusst [29].

Es gibt unterschiedliche Typen von Intermediären, die auch Nachhaltigkeits-Research- oder -Rating-Agenturen, ESG-Research oder Responsible Investment Research (RIR) Groups genannt werden. Spezialisierte Anbieter sind z. B. CDP, RobecoSAM, MSCI und ISS-Oekom [30-32]. Dies sind meist profitorientierte Unternehmen. Eine Sonderrolle hat die Nichtregierungsorganisation CDP. Mittlerweile haben auch klassische Finanz-Research-Agenturen Nachhaltigkeitsinformationen in ihr Portfolio aufgenommen. Intermediäre können selbstständige Organisationen sein oder als interner Teilbereich von Finanz-Intermediären auftreten, die ein Portfolio nachhaltiger Kapitalanlagen managen [29, 33].

3.3 IT-Systeme zur Stakeholder-orientierten Nachhaltigkeitskommunikation

IT-Systeme unterstützen Unternehmen in einer zielgerechten und effizienten Nachhaltigkeitskommunikation aus Nachrichten-Perspektive. Existierende Nachhaltigkeitsmanagement-Software für Unternehmen verfügt teilweise über Reporting-Module, die die Daten nach verbreiteten Standards ausgeben, bspw. SAP SuPM (SAP), SoFi (thinkstep) und Enablon publisher (Enablon). Mehrere Autoren heben das Potenzial des Internets für die Nachhaltigkeitskommunikation von Unternehmen hervor. So werden zielgruppenspezifisch Tiefe und Umfang von Informationen variiert [34]. Haßler [35] beschreibt Chancen und Herausforderungen für Intermediäre, die durch Nachhaltigkeitsberichterstattung im Internet entstehen. Aus Sicht von Kommunikationsagenturen betonen Behrens und Winter [34] die Möglichkeit für den Dialog mit Stakeholdern. Isenmann, Marx Gómez und Süpke [36-37] zeigen das Potenzial von individualisierter, zielgruppengerechter und automatisierter Nachhaltigkeitsberichterstattung mit Hilfe eines Software-Werkzeugs. Auch Solsbach et al. [38] untersuchen serviceorientierte und maßgeschneiderte Nachhaltigkeitsberichterstattung und den Dialog in einem betrieblichen Umweltinformationssystem. Solche IT-Systeme unterstützen den Abschied von der einseitigen und Stakeholder-übergreifenden Kommunikation. Allerdings werden sie bisher in der Praxis nicht weiterverfolgt. Speziell für die Zielgruppe der Investoren birgt die individualisierte Ausgabe von Nachhaltigkeitsinformationen großes Potenzial, weil Investoren ein heterogenes Informationsbedürfnis haben. Als Austauschformat eignet sich dafür XBRL, da es bereits in der digitalen Finanzberichterstattung verbreitet ist [39] und für die Nachhaltigkeitsberichterstattung nach GRI vorliegt [40]. Für die ökologische Dimension der Nachhaltigkeit untersucht eine aktuelle Forschungskooperation der Universität Oldenburg und der Volkswagen AG die prototypische Realisierung einer Unternehmensplattform: „Environmental Communication Data Platform for the Financial Community“ (eco4fin).

4 Forschungsdesign

4.1 Herleitung der Forschungsfrage

Immer mehr Investoren legen ihr Kapital nachhaltig an und nutzen Nachhaltigkeitsinformationen über Unternehmen für die Risikoabschätzung ihrer Investitionsentscheidungen. Diese Nachhaltigkeitsinformationen gelangen über Kommunikationswege zur Stakeholdergruppe der Investoren. So stellen Unternehmen in der Praxis etwa relevante Nachhaltigkeitsinformationen gebündelt für die Stakeholder bereit. Darüber hinaus haben sich Intermediäre etabliert, die die Kommunikation zwischen Unternehmen und Investoren vermitteln. Aus Nachrichten-Perspektive bergen IT-Systeme das Potential, die Nachhaltigkeitskommunikation weiterzuentwickeln, indem sie Stakeholder-orientiert und effizient maßgeschneiderte Informationen ausgeben. Im Vorfeld der Entwicklung solcher IT-Systeme bedarf es der Kenntnis, wie die Nachhaltigkeitskommunikation am Kapitalmarkt aktuell gestaltet ist.

4.2 Methodisches Vorgehen

Die vorliegende Untersuchung analysiert literaturbasiert den Status quo der Nachhaltigkeitskommunikation am Kapitalmarkt. Zunächst wird mit Hilfe einer Literaturarbeit das Phänomen der Nachhaltigkeitskommunikation am Kapitalmarkt beschrieben. So werden Strukturen herausgearbeitet, die in ein strategisches Prozessmodell einfließen. Dieses Modell dient dem Überblick über die Interaktion der beteiligten Akteure und stellt Formate ihrer Interaktion heraus. Damit dient das Modell einer fachlich groben Erklärung des Prozesses. Der strategisch beschreibende Charakter des Modells erfordert eine fachlich exakte Darstellung mit übersichtlichem Detaillierungsgrad. Auf die Modellierung eventueller Ausnahmen und Restriktionen wird verzichtet. Das Modell bildet mit dem Geltungsanspruch eines Ist-Modells die Nachhaltigkeitskommunikation am Kapitalmarkt als gesamten Prozess im Überblick ab. Die Komplexität des Prozesses wird durch starke Abstraktion reduziert. Schwerpunkte werden durch Zerlegung verdeutlicht und teilweise zusätzlich als Teilprozesse abgebildet. Es wird die Organisationssicht gewählt, weil das Erkenntnisinteresse im Verständnis der Rollen im Nachhaltigkeitskommunikationsprozess am Kapitalmarkt liegt. Die wesentlichen Prozessbeteiligten werden als Pools dargestellt: Investor, Intermediär und Unternehmen. Auf eine weitere Untergliederung wird verzichtet, da die Darstellung einen allgemeinen Überblick geben soll und in der Praxis die Beteiligten innerhalb der Pools variieren. Die fachliche Prozessmodellierung der vorliegenden Arbeit nutzt Business Process Modell and Notification Version 2.0 (BPMN 2.0) des Standardisierungsgremiums Object Management Group [41-42]. Die Spezifikationssprache ist ein weltweit führender De-facto-Standard für Geschäftsprozesse [43]. BPMN 2.0 ist in Wissenschaft, Informatik und im praktischen Geschäftsleben etabliert und ermöglicht die Einbindung von Artefakten, Objekten und Datenspeichern.

5 Ergebnisse

5.1 Literaturarbeit

Mit der Situation der Nachhaltigkeitskommunikation befassen sich in den Wirtschaftswissenschaften unterschiedliche Modelle. Diese behandeln die Kommunikation und Interaktion der Akteure am Markt für nachhaltige Kapitalanlagen [18, 29, 44]. Ein häufig zitiertes Modell der Akteure stammt von Schönheit [44] und zeigt neben dem Wertfluss auch den Informationsfluss [29, 33]. Der Informationsfluss verläuft einerseits direkt an Investoren und andererseits über Nachhaltigkeits-Research-Agenturen und Finanz-Intermediäre. Neuere Modelle zeichnen den Kommunikationsfluss der Unternehmen an Investoren allein über die Intermediäre. So reduzieren Schäfer et al. [18] ihr

Modell auf den Informationsbedarf und stellen Nachhaltigkeits-Research-Agenturen als Informations-Intermediäre dar. Diese versorgen die Investoren entweder direkt mit angefragten Informationen oder stellen wahlweise Zusatzinformationen über Informationsprodukte zur Verfügung [18]. Auch Arnold [29] betont den Zwei-Stufen-Fluss der Kommunikation am SRI-Markt zwischen Unternehmen und Investoren. Auf der mittleren Kommunikationsstufe schließt er neben Nachhaltigkeits-Research-Agenturen auch Finanz-Intermediäre ein. Die vorgestellten Modelle zeigen teilweise begriffliche und inhaltliche Unterschiede, die sich mit unterschiedlichen Forschungsschwerpunkten begründen lassen. Auffällig ist, dass die direkte Kommunikation zwischen Unternehmen und Investoren in den neueren Modellen wegfällt, und so die zentrale Rolle der Intermediäre betont wird.

Hinsichtlich der Nachhaltigkeitskommunikation beschreibt die Fachliteratur ein Kommunikationsproblem zwischen Unternehmen und Investoren. Denn aus Nachrichten-Perspektive greifen Investoren bei der Recherche von Nachhaltigkeitsinformationen weniger auf Primärinformationen zurück, die ihnen Unternehmen direkt zur Verfügung stellen [16, 45, 46]. Häufig verfolgen Unternehmen bei der Informationsverarbeitung und -bereitstellung einen Multi-Stakeholder-Ansatz. Das gilt sowohl für die Zustellungsformen (Nachhaltigkeitsbericht) als auch für die am häufigsten zugrundeliegenden Rahmenwerke wie die der GRI [26]. Speziell auf Investoren ausgerichtet ist bspw. das neu erarbeitete Rahmenwerk für klimarelevante Berichterstattung der Task Force on Climate-related Financial Disclosure (TCFD) [23]. Ältere weitergefasste Rahmenwerke und Standards mit speziellem Investorenbezug finden sich bei Schäfer et al. [18] und Bassen & Senkl [24]. Kapitalmarktorientiert und europaweit bekannt ist der freiwillige Standard für Schlüsselkriterien (ESG-KPI), der von der Deutschen Vereinigung für Finanzanalyse und Asset Management (DVFA) erarbeitet und von der europäischen Dachvereinigung EFFAS übernommen wurde [25]. Trotz der vielen Rahmenwerke und Standards ist die direkte Nachhaltigkeitskommunikation zwischen Unternehmen und Investoren nur ansatzweise vorhanden. Insb. greifen Investoren auf folgendes Material zurück, das sie relativ mühelos für direkte Analysen abrufen können: Nachhaltigkeitsbericht, Informationen vom Internetauftritt des Unternehmens, Lagebericht im Rahmen des Geschäftsberichts und Berichterstattung an investorenunterstützte Initiativen [16]. Dadurch entgehen den Investoren unternehmensrelevante Nachhaltigkeitsinformationen, die an anderen Orten bereitstehen, bspw. in Broschüren oder in einzeln veröffentlichten Umwelterklärungen. Für die Investitionsentscheidung schließlich nutzen Investoren die von Unternehmen bereitgestellten Nachhaltigkeitsinformationen lediglich eingeschränkt bis gar nicht, und kritisieren deren Nachhaltigkeitsberichterstattung [16, 45]. Für diese mangelnde Nutzung von Informationen finden sich in der Literatur mehrere Gründe auf Seiten der Unternehmen und des Kapitalmarkts [16, 46]. Die zugrundeliegenden Ursachen dafür liegen meist im mangelnden Verständnis der jeweils anderen Interessen und Grenzen [16]. Beheben könnte dies ein verbesserter Dialog zu Nachhaltigkeitsthemen [45]. Bspw. initiieren Investor-Relations-Abteilungen der Unternehmen proaktiv Investoren-Veranstaltungen (sog. Roadshows), Telefonkonferenzen und Einzelgespräche [18].

Bisher haben Intermediäre eine Brückenfunktion eingenommen. Sie prägen die Nachhaltigkeitskommunikation sowohl technisch als auch inhaltlich, denn Investoren nutzen vorrangig Primär- und Sekundärinformationen, die Intermediäre bereitstellen [16, 30-31, 45]. Auf der einen Seite sammeln Intermediäre in vorgefertigten Fragebögen Primärinformationen und stellen diese unbewertet den Investoren bereit (z. B. CDP, RobecoSAM). Auf der anderen Seite tragen sie die Informationen in einer eigenen Datenbasis zusammen. Auf der Grundlage dieser Informationsbasis generieren sie durch Analysieren und Bewerten Sekundärinformationen als eigene Produkte, die sie – individualisiert oder standardisiert – an Investoren weitergeben. Die Produkte der Intermediäre reichen von unbewerteten Informationsprofilen bis hin zu abgeschlossenen Bewertungen der Nachhaltigkeitsperformance des betrachteten Unternehmens. Bspw. veröffentlichen viele Intermediäre Nachhaltigkeits-Rating-Produkte. Dazu zählen Nachhaltigkeits-Indizes, -Rankings und -Ratings [32], wie etwa der Dow Jones Sustainability Index (DJSI). Es existieren viele unterschiedliche Research-Agenturen und Rating-Produkte [31-32]. In den letzten Jahren wurden jedoch Tendenzen der Konsolidierung in der Nachhaltigkeits-Rating-Industrie erkennbar [30].

Intermediäre sind auf Nachhaltigkeitsinformationen von Unternehmen angewiesen, um zuverlässige Ergebnisse zu liefern [19]. Informationsgrundlage und -analyse können auf frei verfügbaren Unternehmensinformationen oder auf zusätzlich beschafften Informationen basieren. Letztere stammen aus Ad-hoc-Abfragen bei Unternehmen oder aus Informationsdatenbanken Dritter [16, 33]. Insb. Informationen über börsennotierte Großunternehmen führen Intermediäre gern im Portfolio, weil Investoren diese häufig nachfragen [16, 18]. Informationen werden von Analysten ausgewertet, die über das notwendige Fachwissen verfügen. Häufig fordert die Fachliteratur eine bessere Qualität und Verfügbarkeit von Informationen für Intermediäre. Schritte in diese Richtung sind erkennbar. So etwa, weil Unternehmen davon ausgehen, dass sie durch zielgruppengerechte Kommunikation und Informationsbereitstellung Ratingergebnisse beeinflussen können [29].

Ohne den Blick auf Intermediäre finden sich in der Literatur unterschiedliche Ansätze, um das Kommunikationsproblem zwischen Unternehmen und Investoren zu lösen. Laut Schäfer et al. [18] bedarf es dazu keiner neuen Rahmenwerke und Standards. Stattdessen schlagen sie ein kapitalmarktorientiertes Leistungsindikatorsystem auf Grundlage bestehender Richtlinien vor. Sullivan [16] hingegen argumentiert, dass sowohl Indikatorsysteme als auch bessere Standards das Kommunikationsproblem nicht lösen könnten, weil Investoren bestrebt seien, sich voneinander abzuheben. Auch weist er darauf hin, dass Unternehmen nur begrenzte Ressourcen hätten, um die von Investoren gewünschten Informationen bereitzustellen [16]. Eine solche Bereitstellung haben Unternehmen jedoch i. d. R. nicht etabliert. Für eine bessere direkte Ansprache von Investoren können Unternehmen ihre Nachhaltigkeitskommunikation im Hinblick auf deren Nutzen anpassen [10, 16, 18]. Arnold [29] schlägt eine eigene zielgerichtete Nachhaltigkeitskommunikation für den Kapitalmarkt vor. Dazu werden aber die inhaltlichen und technischen Anforderungen von Investoren in der Literatur nur indirekt angesprochen [45-46] und nicht strukturiert erhoben. Einige Forschungsarbeiten schlagen einen maßgeschneiderten, dialog- und internetbasierten Kommunikationsansatz vor [18, 40, 38, 47, 48]. Dies entspricht einem Wechsel von der einseitigen (Push-) zur zweiseitigen (Pull-) Kommunikation [49]. Auch wird das Zusammenlegen der Finanz- und Nachhaltigkeitsberichterstattung diskutiert [46], was in der Praxis teilweise schon umgesetzt wird. Bisher lassen sich jedoch keine überzeugenden Hinweise auf eine Verbesserung des beschriebenen Kommunikationsproblems in der Praxis finden. In dieser Situation soll das Handbuch von Sullivan [16] Investoren im Umgang mit dem derzeit verfügbaren Datenmaterial der Unternehmen unterstützen.

5.2 Prozessmodell aus Nachrichten-Perspektive

Abbildung 1 zeigt das Ist-Modell der Nachhaltigkeitskommunikation zwischen den Kapitelmarkt-Akteuren im Überblick. Zwei Ausschnitten stellen die aufgegliederten Subprozesse dar: Abbildung 2 zeigt die Informationsbeschaffung des Intermediärs, Abbildung 3 den Informationsaustausch zwischen Investor und Intermediär.

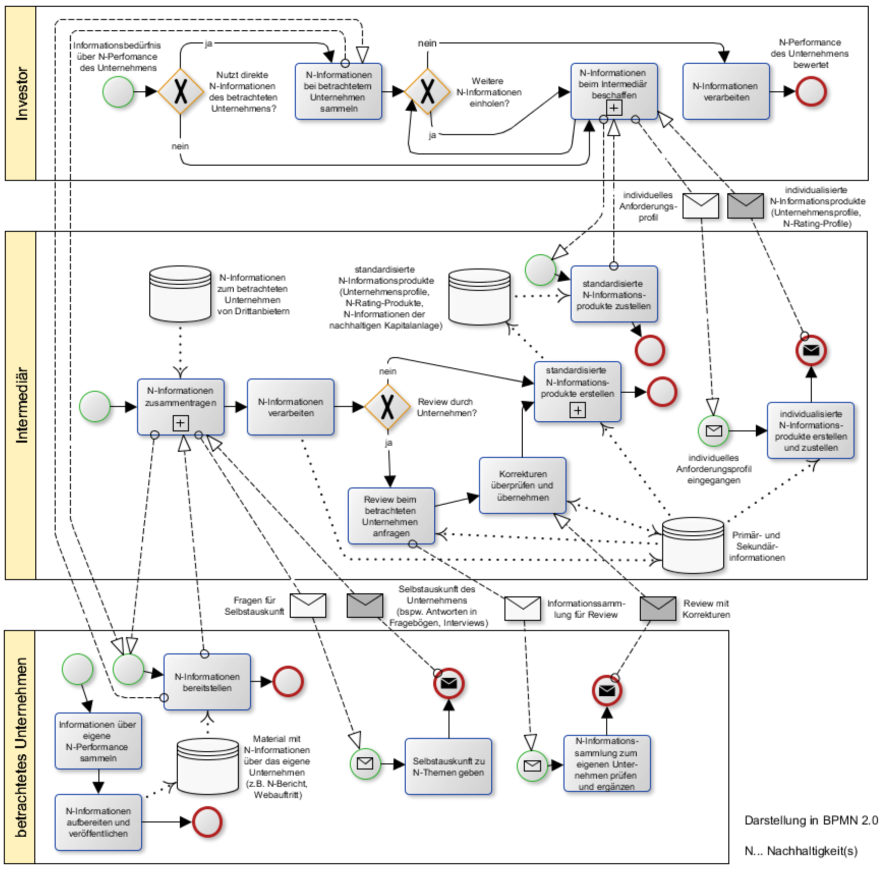


Abbildung 1. Nachhaltigkeitskommunikation am Kapitalmarkt

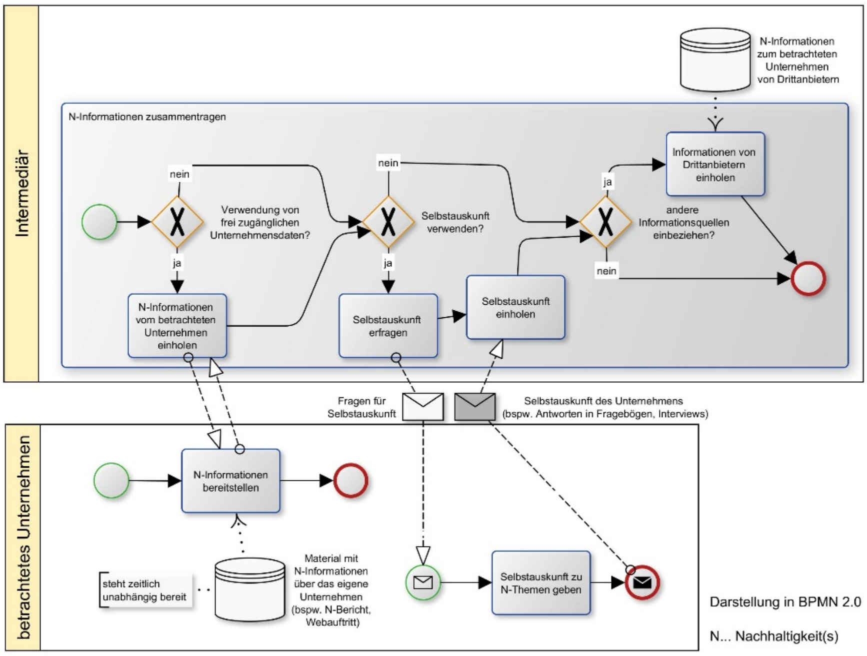


Abbildung 2. Informationsbeschaffung des Intermediär

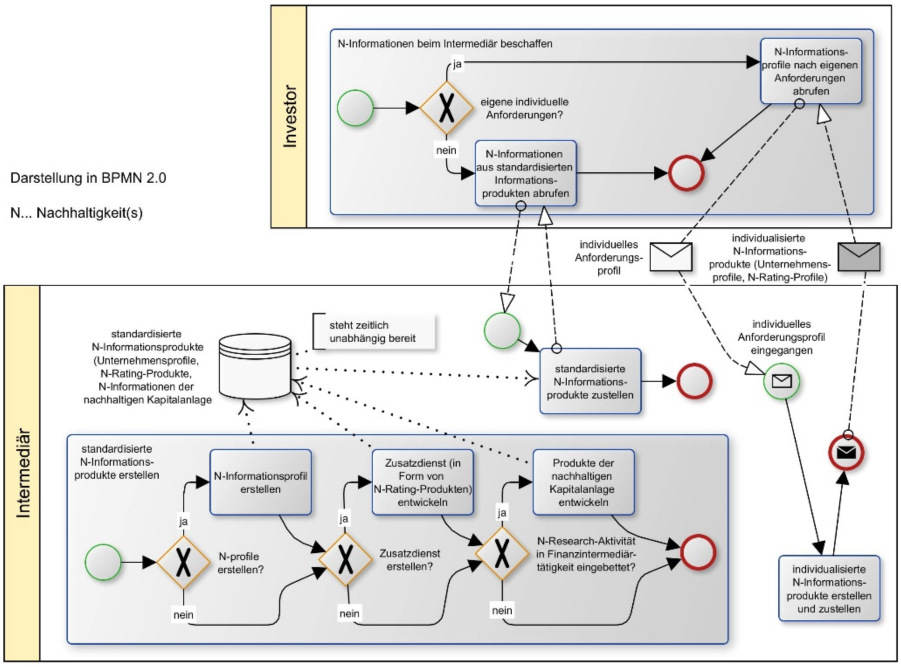


Abbildung 3. Informationsaustausch zwischen Investor und Intermediär

Initiiert wird der Prozess der Nachhaltigkeitskommunikation – mit Blick auf den Kapitalmarkt – durch das Informationsbedürfnis des Investors. Dieser möchte die Nachhaltigkeitsperformance des Unternehmens bewerten. Gegebenenfalls sammelt er zu diesem Zweck Informationen direkt vom Unternehmen. Zusätzlich kann er sich weitere Informationen von einem oder mehreren Intermediären beschaffen (Subprozess: *N-Informationen beim Intermediär beschaffen*). Die gesammelten Informationen verarbeitet der Investor und bewertet die Nachhaltigkeitsperformance.

Ausgelöst vom Informationsbedürfnis des Investors sammelt ein Unternehmen intern Informationen über die eigene Nachhaltigkeitsperformance und bereitet diese auf. Schließlich veröffentlicht es diese, bspw. als Nachhaltigkeitsbericht oder in Form eines Webauftritts. Unternehmensseitig stehen damit Nachhaltigkeitsinformationen in aufbereiteter Form öffentlich zur Verfügung.

Der Intermediär trägt Nachhaltigkeitsinformationen zum betrachteten Unternehmen zusammen (Subprozess: *N-Informationen zusammentragen*), verarbeitet diese und sammelt sie. Räumt der Intermediär dem Unternehmen anschließend die Möglichkeit zum Review ein, so erhält er auf Anfrage vom Unternehmen eine Überprüfung und Ergänzung seiner Informationssammlung. Nach Überprüfung übernimmt der Intermediär die Korrekturen. Auf dieser Basis kann er standardisierte Nachhaltigkeitsinformationsprodukte erstellen (Subprozess: *standardisierte N-Informationsprodukte erstellen*). Hierbei bereitet der Intermediär seine Informationssammlung zum standardisierten Abruf vor. Zusätzlich hat er die Möglichkeit, diese Sammlung für individualisierte Nachhaltigkeitsinformationsprodukte zu nutzen, die er auf Anfrage eines Investors erstellt.

Abbildung 2 zeigt die Informationsbeschaffung des Intermediärs detaillierter als in der Übersicht. Der Intermediär hat unterschiedliche und miteinander kombinierbare Möglichkeiten, Nachhaltigkeitsinformationen über das betrachtete Unternehmen zu sammeln. Der Intermediär kann dabei, ebenso wie der Investor, auf frei zugängliches Informationsmaterial des Unternehmens zurückgreifen. Weiterhin kann er auf Anfrage eine Selbstauskunft vom Unternehmen erhalten. Zusätzlich können Nachhaltigkeitsinformationen von Dritten hinzugezogen werden, die i. d. R. nicht frei zugänglich sind.

Die Übersichtsdarstellung in Abbildung 1 verdeutlicht den direkten Kontakt von Intermediär und Unternehmen im Prozess der Nachhaltigkeitskommunikation. Der Intermediär kann den Kontakt herstellen, indem er das Unternehmen zu einer Selbstauskunft auffordert, welche i. d. R. einen Fragebogen oder ein Interview umfasst. Darüber hinaus kann der Intermediär dem Unternehmen ein Review anbieten, das die gesammelten Nachhaltigkeitsinformationen über dieses Unternehmen beinhaltet. Das Unternehmen gibt das Review mit Korrekturen und Ergänzungen zurück.

Abbildung 3 veranschaulicht den Informationsaustausch zwischen Investor und Intermediär über standardisierte und individuelle Informationsprodukte. Für den Investor zeigt Abbildung 3 den aufgegliederten Subprozess *N-Informationen beim Intermediär beschaffen*. Zur Auswahl stehen mehrere standardisierte Produkte. Entscheidet sich der Investor für individualisierte Produkte, so ist dafür im Vorfeld die Übermittlung eines individuellen Anforderungsprofils an den Intermediär notwendig.

Für den Intermediär veranschaulicht Abbildung 3 den aufgegliederten Subprozess *standardisierte N-Informationsprodukte erstellen*. Daneben ist die Erstellung individualisierter Informationsprodukte abgebildet, die der Intermediär gemäß den Anforderungen des Investors realisiert. Individualisierte Produkte sind bspw. Unternehmensprofile und Nachhaltigkeitsratingprofile. Standardisierte Produkte kann der Intermediär bereits im Vorfeld einer Investoren-Anfrage, auf Grundlage seiner Nachhaltigkeitsinformationssammlung, erstellen. Dabei bieten sich dem Intermediär mehrere Möglichkeiten: (1) Er kann ein standardisiertes Informationsprofil über das betrachtete Unternehmen erstellen oder (2) er kann Zusatzdienste wie standardisierte Nachhaltigkeitsratingprodukte (Indizes, Rankings, Ratings) anbieten. (3) Er kann – sofern er Teil eines Finanz-Intermediärs ist – die Nachhaltigkeitsinformationen ergänzend zur internen Finanzanalyse nutzen und Produkte zur nachhaltigen Kapitalanlage entwickeln. Die Informationen aus diesen Anlageprodukten werden ebenfalls als standardisierte Informationsprodukte eingeordnet.

6 Diskussion, Fazit und Ausblick

Mit der Modellierung ist eine aktuelle Übersicht gelungen, die die Möglichkeiten der Nachhaltigkeitskommunikation am Kapitalmarkt zwischen den Akteuren darstellt. Das Modell zeigt, dass Investoren auf die Dienste von Intermediären zugreifen – entweder als alleinige Informationsgrundlage oder als Ergänzung zu selbst recherchierten Unternehmensinformationen. Allerdings ist aus der Forschung bekannt, dass viele Investoren die von Unternehmen bereitgestellten Nachhaltigkeitsinformationen nicht für ihre Investitionsentscheidungen verwenden. Daraus und aus der Tatsache, dass viele Investoren dennoch nachhaltige Kapitalanlagen tätigen, wird auf die zentrale Position der Intermediäre geschlossen. Bildhaft kann dieses Phänomen als Nadelöhr im Informationsfluss bezeichnet werden.

Folgende Limitationen liegen vor. Zum einen basiert das Modell auf einer Literaturrecherche und wurde nicht empirisch validiert, sodass Lücken oder Verzerrungen im Vergleich zur Diskurswelt möglich sind. Zum anderen führen methodische Grenzen der Prozessmodellierung dazu, dass das Modell keine Aussage über die praktische Relevanz der aufgezeigten Möglichkeiten geben kann. Außerdem macht das Modell keine Aussagen über Inhalte und Formate der Nachhaltigkeitskommunikation. Auch die Heterogenität und Relevanz von Investoren und Intermediären wird vernachlässigt. Jedoch deuten die Artefakte auf individuell verschiedene Anforderungen hin. Darüber hinaus vernachlässigt das Modell die Verständigung, die aus Sprechakt-Perspektive Voraussetzung für eine erfolgreiche Kommunikation auf Nachrichtenebene ist. Insb. zum Geltungsanspruch der Wahrhaftigkeit ist möglicherweise ein Diskurs notwendig, da nicht per se davon auszugehen ist, dass die Intentionen der Akteure wechselseitig anerkannt sind; d. h. die Glaubwürdigkeit von Unternehmen (z. B. Greenwashing), Investoren (z. B. ehrliches Interesse) und Intermediären (z. B. Intransparenz).

Aktuell nutzen vor allem Intermediäre in ihrer Rolle als Nadelöhr die Möglichkeit, den Faden der Nachhaltigkeitskommunikation am Kapitalmarkt zu führen. Dies kann zur Konzentration der Deutungshoheit führen. Dem entgegen steht der Wettbewerb unter den Intermediären, die sich durch unterschiedliches Nachhaltigkeitsverständnis und Deutungspluralität auszeichnen. Ein Lösungsansatz besteht in der

Dezentralisierung der Informationsbereitstellung. Das meint aus Nachrichten-Perspektive eine Stakeholder-orientierte Bereitstellung von Primärinformationen von Unternehmen direkt an Investoren. Jedoch trägt eine neue Infrastruktur nicht dazu bei, ein möglicherweise bestehendes Verständigungsproblem im Kommunikationsprozess zu lösen. Daher unterstützt aus Sprechakt-Perspektive der Diskurs über bestehende Verständigungsprobleme (z. B. Glaubwürdigkeit der Akteure) die Weiterentwicklung einer zielführenden Nachhaltigkeitskommunikation. Durch eine aus diesen beiden Kommunikationsperspektiven verbesserte Nachhaltigkeitskommunikation entsteht eine solide Grundlage, die Investoren und Intermediäre größere Sicherheit bei der Bewertung der Nachhaltigkeitsperformance von Unternehmen verleihen kann. Ebenso entstehen dadurch für Unternehmen Vorteile, wie Effizienz durch Automatisierung und Ableiten von Abfrage-Bedarfsanalysen. Die zentrale Rolle der Intermediäre könnte durch die Umsetzung von alternativen Strukturen der Informationsbereitstellung aus Nachrichten-Perspektive entfallen. Aus der Sprechakt-Perspektive wird jedoch zusätzlich deutlich, dass eine zentrale Rolle der Intermediäre unter Umständen weiterbesteht, da diese zur Verständigung (Geltungsanspruch Verständlichkeit) beitragen. Denn nur ein Teil der Investoren hat das Fachwissen zur Bewertung der Informationen. Der andere Teil bleibt auf die Expertise von Intermediären angewiesen.

Die Ergebnisse schaffen eine Basis für weitere Forschung zur Nachhaltigkeitskommunikation am Kapitalmarkt im Bereich der Wirtschaftsinformatik. Forschungsaktivitäten sind zu betrieblichen umwelt- und nachhaltigkeitsbezogenen Informations- und Kommunikationssystemen denkbar. Praktikern bietet das Modell einen Ansatz für die gezielte Weiterentwicklung der Nachhaltigkeitskommunikation im eigenen Unternehmen. Der vorliegende Beitrag zeigt, dass IT-Systeme dazu beitragen können, die Nachhaltigkeitskommunikation von Unternehmen zielgerecht und effizient auf Investoren und Intermediäre auszurichten. An das aufgezeigte Spannungsfeld knüpft die Forschungskooperation „Environmental Communication Data Platform for the Financial Community“ (eco4fin) an. Das Projekt untersucht den Einsatz einer Plattform zur Umweltkommunikation von Großunternehmen mit Investoren und Intermediären.

Disclaimer. Ergebnisse, Meinungen und Schlüsse dieser Veröffentlichung sind nicht notwendigerweise die der Volkswagen Aktiengesellschaft.

Referenzen

1. GSIA: Global Sustainable Investment Review 2016. Report, GSIA (2017)
2. HLEC: Financing a Sustainable European Economy. Final Report, HLEG (2018)
3. Weber, J., Bramsemann, U., Heineke, C., Hirsch, B.: Kapitalmarktkommunikation. In: Wertorientierte Unternehmenssteuerung. pp. 233–268. Springer, Wiesbaden (2017)
4. Beck, K.: Kommunikationswissenschaft /. UVK-Verl.-Ges., (2015).
5. Abts, D., Mülder, W.: Grundkurs Wirtschaftsinformatik: eine kompakte und praxisorientierte Einführung. Springer Vieweg, Wiesbaden (2017).
6. Flores, F., Graves, M., Hartfield, B., Winograd, T.: Computer Systems and the Design of Organizational Interaction. ACM Transactions on Information Systems. 6, 153–172 (1988).
7. Burkart, R.: Kommunikationswissenschaft. Böhlau, Wien (2002).

8. Winograd, T.: A Language/Action Perspective on the Design of Cooperative Work. Human–Computer Interaction. 3, 3–30 (1987).
9. von Wallis, M., Klein, C.: Ethical Requirement and Financial Interest: A Literature Review on Socially Responsible Investing. Business Research. 8, 61–98 (2015)
10. Eccles, R.G., Serafeim, G., Krzus, M.P.: Market Interest in Nonfinancial Information. Journal of Applied Corporate Finance. 23, 113–127 (2011)
11. PRI, About the PRI, https://www.unpri.org/pri/about-the-pri (Aufgerufen: 29.08.2018)
12. CDP, Become an Investor Member, https://www.cdp.net/en/investor/signatories-and-members (Aufgerufen: 13.09.2018)
13. Cengiz, C.-B., Braun, D., von Nitzsch, R.: Alpha-Vehikel oder Preis für das gute Gewissen: Eine Performanceanalyse ethischer Investments. Corporate Finance biz. 04, 263–271 (2010)
14. Schäfer, H.: Sozial-ökologische Ratings am Kapitalmarkt. Hans-Böckler-Stiftung Nr. 84, Düsseldorf (2003)
15. Friede, G., Busch, T., Bassen, A.: ESG and Financial Performance: Aggregated Evidence from Tore than 2000 Empirical Studies. Journal of Sustainable Finance & Investment. 5, 210–233 (2015)
16. Sullivan, R.: Valuing Corporate Responsibility: How Do Investors Really Use Corporate Responsibility Information? Routledge, London (2011)
17. HLEC: Financing a Sustainable European Economy. Interim Report 07/ 2017, HLEG (2017)
18. Schäfer, H., Kröner, F., Seeberg, B.: Zielgruppengerechte Nachhaltigkeitsberichterstattung mit Hilfe des Socially Responsible Investment (SRI)-Factsheets. Controlling. 23, 466–472 (2011)
19. Schäfer, D.H.: Ausschlusskriterien in der nachhaltigen Geldanlage: Eine ökonomische Analyse. Forschungsbericht Nr. 01/2014, Universität Stuttgart (2014)
20. Sandberg, J., Juravle, C., Hedesström, T.M., Hamilton, I.: The Heterogeneity of Socially Responsible Investment. Journal of Business Ethics. 87, 519 (2008)
21. Schäfer, H.: Verantwortliches Investieren. In: Corporate Social Responsibility auf dem Finanzmarkt. pp. 64–80. VS, Wiesbaden (2009)
22. Siew, R.Y.J.: A Review of Corporate Sustainability Reporting Tools (SRTs). Journal of Environmental Management. 164, 180–195 (2015)
23. TCFD: Recommendations of the Task Force on Climate-related Financial Disclosures. Final Report, TCFD (2017)
24. Bassen, A., Senkl, D.: Leistungsindikatoren nachhaltiger Unternehmensführung. Controlling: Zeitschrift für erfolgsorientierte Unternehmenssteuerung. 22, 256–261 (2010)
25. EFFAS und DVFA: KPIs für ESG 3.0. A Guideline for the Integration of ESG into Financial Analysis an Corporate Valuation (Version 3.0). EFFAS (2010)
26. KPMG: The KPMG Survey of Corporate Responsibility Reporting 2017. (2017). https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2017/10/kpmg-survey-of-corporate-responsibility-reporting-2017.pdf (Aufgerufen: 05.10.2018)
27. Lehmann, C., Ruziczka, D., Pöhlmann, T., Huck-Sandhu, S.: Von Skeptikern, Sympathisanten und Zuschauern: Stakeholder-Typen und ihre Erwartungen an die Nachhaltigkeitsberichterstattung. NachhaltigkeitsManagementForum. (2018).
28. Friedrich, C.: Nachhaltigkeitskommunikation als Motor für die Weiterentwicklung von Stakeholder-Dialogen. uwf. 21, 39–43 (2013).
29. Arnold, J.: CSR-Kommunikation am nachhaltigen Kapitalmarkt. uwf. 19, 229–236 (2011)
30. Avetisyan, E., Hockerts, K.: The Consolidation of the ESG Rating Industry as an Enactment of Institutional Retrogression. Business Strategy and the Environment. 26, 316–330 (2017)
31. Schäfer, H., Zenker, J., Beer, J., Fernandes, P.: Who is Who in Corporate Social Responsibility Rating, Research Report. Bertelsmann-Stiftung, Stuttgart (2006)

32. GISR, http://ratesustainability.org (Aufgerufen: 11.4.2018)
33. Berg, R.: Legitimation durch Informationsintermediäre mit Nachhaltigkeitsbezug. Springer, Wiesbaden (2018)
34. Behrens, B., Winter, M.: Anforderungen an ein Stakeholder-Reporting aus Sicht der Kommunikationsagentur. In: Isenmann, R. and Marx Gómez, J. (eds.) Internetbasierte Nachhaltigkeitsberichterstattung. pp. 241–252. Erich Schmidt, Berlin (2008)
35. Haßler, R.: Internetgestütze Nachhaltigkeitsberichterstattung: Bedeutung für das Nachhaltigkeits-Rating. In: Isenmann, R. and Marx Gómez, J. (eds.) Internetbasierte Nachhaltigkeitsberichterstattung. pp. 231–240. Erich Schmidt, Berlin (2008)
36. Isenmann, R., Marx Gómez, J.: Software-Werkzeug zur automatisierten und maßgeschneiderten Nachhaltigkeitsberichterstattung. In: Isenmann, R. and Marx Gómez, J. (eds.) Internetbasierte Nachhaltigkeitsberichterstattung. pp. 337–354. Erich Schmidt, Berlin (2008)
37. Isenmann, R., Marx Gómez, J., Süpke, D.: Making Stakeholder Dialogue for Sustainability Issues Happen: Benefits, Reference Architecture and Pilot Implementation for Automated Sustainability Reporting a La Carte. In: 44th Hawaii International Conference on System Sciences (HICSS). pp. 1–11. IEEE Press, Koloa (2011)
38. Solsbach, A., Lipnitskaya, S., van Vliet, S.: Internetbasierte Nachhaltigkeitsberichterstattung im Kontext des Umwelt-, Energie- und Ressourcenmanagements mit BUIS der nächsten Generation. In: Marx Gómez, J., Lang, C., and Wohlgemuth, V. (eds.) IT-gestütztes Ressourcen- und Energiemanagement. pp. 421–431. Springer, Berlin (2013)
39. Kesselmeyer, B., Leibfried, P.: Standardisierung der Berichterstattung durch XBRL. In: Keuper, F., Vocelka, A., and Häfner, M. (eds.) Die moderne Finanzfunktion: Strategien, Organisation und Prozesse. pp. 279–299. Gabler, Wiesbaden (2008)
40. Solsbach, A.: Document Engineering als Ansatz für eine überbetriebliche Nachhaltigkeitsberichterstattung. Shaker, Aachen (2015)
41. Object Management Group (OMG): Business Process Model and Notation (BPMN): Version 2.0. OMG Document Number: formal/2011-01-04 (2011)
42. Object Management Group (OMG): BPMN 2.0 by Example, Version 1.0 (non-normative). OMG Document Number: dtc/2010-06-02. (2010)
43. Chinosi, M., Trombetta, A.: BPMN: An Introduction to the Standard. Computer Standards & Interfaces. 34, 124–134 (2012)
44. Schönheit, I.: Markttransparenz im Socially Responsible Investment: Konsequenzen für eine nachhaltige Erwachsenenbildung. Lang, Frankfurt a. M. (2005)
45. Sullivan, R., Gouldson, A.: Does Voluntary Carbon Reporting Meet Investors' Needs? Journal of Cleaner Production. 36, 60–67 (2012)
46. Frank, R.: Integrierte Berichte: Gehört dem Integrated Reporting die Zukunft? In: Schulz, T. and Bergius, S. (eds.) CSR und Finance. pp. 237–250. Springer, Berlin (2014)
47. Süpke, D., Gómez, J.M., Isenmann, R.: Stakeholder Interaction in Sustainability Reporting with Web 2.0. In: Athanasiadis, I.N., Rizzoli, A.E., Mitkas, P.A., and Marx Gómez, J. (eds.) Information Technologies in Environmental Engineering. pp. 387–398. Springer, Berlin (2009)
48. Isenmann, R., Marx Gómez, J.: Software-Werkzeug zur automatisierten und maßgeschneiderten Nachhaltigkeitsberichterstatung. In: Isenmann, R. and Marx Gómez, J. (eds.) Internetbasierte Nachhaltigkeitsberichterstattung. pp. 337–354. Erich Schmidt, Berlin (2008)
49. Bruhn, M.: Integrierte Unternehmens- und Markenkommunikation: Strategische Planung und operative Umsetzung. Schäffer-Poeschel, Stuttgart (2014)

Potentiale und Herausforderungen der Materialflusskostenrechnung

Felix Hemke[1], Anna Lütje[1,2], Hans-Knud Arndt[3], Volker Wohlgemuth[1]

[1] Hochschule für Technik und Wirtschaft Berlin, Wilhelminenhofstr. 75A, 12459 Berlin, Germany
{Felix.Hemke, Anna.Luetje, Volker.Wohlgemuth}@htw-berlin.de
[2] Leuphana Universität Lüneburg, Germany
Anna.Luetje@htw-berlin.de
[3] Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, 39016 Magdeburg, Germany
Hans-Knud.Arndt@iti.cs.uni-magdeburg.de

Abstract. The article gives a description of the method of material flow cost accounting. Secondly general obstacles and capabilities are described. The gathering of the necessary data is a serious problem for applying companies. A series of software requirements, with the potential to reduce the problem, were determined through a usability study and are presented in the last part. The implementation of these requirements can help to diminish the obstacles and help with the diffusion of the method.

Keywords: Material Flow Cost Accounting, Software Requirements

1 Einführung

Die Materialflusskostenrechnung (engl. Material Flow Cost Accounting, kurz MFCA-Methode) zielt darauf ab, die Kosten, die aus Verlusten während der Erzeugung von Produkten in der Prozess- und Fertigungsindustrie entstehen, aufzudecken. Damit bilden die Ergebnisse eine Entscheidungsgrundlage für eine Reduzierung der Materialkosten. Ungewünschte Nebenprodukte im Produktionsprozess lassen sich lokalisieren und minimieren. Damit ist sie geeignet, eine Verringerung von schädlichen unternehmerischen Umweltwirkungen bei gleichzeitiger Erhöhung der ökonomischen Effizienz zu realisieren. Die klassische Kostenkalkulation eines Produktes basiert darauf, dass alle in die Produktion fließenden Rohstoffe letztlich im Produkt landen und in die Preisgestaltung einbezogen werden (Vgl. [1] S. 3-5, [2] S. 3-6). Dies ist insofern konsequent, dass das produzierende Unternehmen diese Kosten in der Produktion hatte und in den Endpreis des Produktes einpreisen muss, um keine Verluste zu machen. Andererseits verhindert es eine zielgerichtete Reduzierung von Ressourcenverschwendung und vermeidbarer Kosten. Die Menge an Rohstoffen, die als Abfall nicht im Produkt landen, wird dabei vernachlässigt. Jedoch macht nur die Kenntnis dieser Mengen und Kosten, eine Reduzierung des Abfalls möglich. Der

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Vorteil der Materialflusskostenrechnung gegenüber der Stoffstromanalyse liegt in der einfacheren Vergleichbarkeit, da das Ergebnis in einer einzelnen monetären Kennzahl (Kosten der Materialverluste) angegeben wird, während die Stoffstromanalyse in der Regel eine Ökobilanz mit der Aufstellung von Umweltwirkungen zum Ergebnis hat, die schwieriger zu vergleichen sind. Das Ziel des Beitrags ist es, die Methode besser nutzbar zu machen, indem diese durch eine anwenderfreundliche Software unterstützt wird.

Die Forschungsfragen des vorliegenden Beitrags sind:

- Welche Vorteile bietet die Methode?
- Welche Schwierigkeiten verhindern eine weitere Ausbreitung?
- Welche technischen Hilfsmittel können zu einer Reduzierung der Schwierigkeiten beitragen?
- Wie lassen sich die Ergebnisse einer Materialflusskostenrechnung nutzen, um damit Kosten und negative Umweltauswirkungen zu reduzieren?

Zunächst erfolgt eine grundlegende Beschreibung zum Verständnis der Methode, da dieses wichtig ist, um Softwareanforderungen zur Methodenumsetzung zu verstehen. Anschließend wird die bisherige Verbreitung der Methode dargestellt. Eine Reihe von Hindernissen, die eine stärkere Akzeptanz bisher verhindern, kommt im darauf folgenden Abschnitt zur Sprache. Die Hindernisse werden schließlich aufgegriffen, um eine Reihe von Anforderungen an eine Software zu definieren, die das Potential besitzen, die Hindernisse zu reduzieren. Den letzten Abschnitt bilden ein Fazit und ein Ausblick auf die weitere Entwicklung.

2 Vorgehen

Die Idee zu dem Beitrag entstand durch Beobachtungen bei Unternehmen, die sich für die Reduzierung der Materialverluste interessieren, sich jedoch, aufgrund bestehender Probleme, vor der Umsetzung der Methode scheuen. Die Befragung von Unternehmensvertretern ließ den Schluss zu, dass die Softwareunterstützung die Methode nicht wesentlich vereinfacht. Anschließend wurde mittels systematischer Literaturrecherche nach Quellen gesucht, die einen tieferen Einblick in die Methodik zulassen. Eine Betrachtung der bestehenden Softwarelösungen, führte dann zu einer Reihe von Fragen, die sich mittels einer Benutzerbefragung beantworten lassen. Aus diesem Grund wurde ein typisches Vorgehen bei der Methodenanwendung ausgearbeitet und Benutzer im Rahmen einer Usability-Studie mit der Bearbeitung beauftragt.

3 Methodenbeschreibung

Die Durchführung einer Materialflusskostenrechnung setzt zunächst die Festlegung eines Untersuchungsobjekts voraus (z.B. das gesamte Unternehmen oder ein bestimmtes Produkt (Vgl. [3], S. 30)). Für das gewählte Betrachtungsobjekt, werden die generellen Stoffströme sowie die darin enthaltenen Materialverluste quantitativ erfasst. Zu diesen gehören alle Stoffe und Energien, die nicht direkt ins Endprodukt fließen, z.B.:

- Ungewünschte Nebenprodukte (z.B. Späne, Schnittreste, Stanzstreifen, Rückstände) (Vgl. [4], S. 87)
- Verpackungsabfälle von Vorprodukten, Roh-, Hilfs- und Betriebsstoffen
- Abfälle durch unsachgemäße Lagerung (z.B. durch Verderb, Verfall) (Vgl. [4], S. 87)
- ungenutzte Materialien (z.B. Körnerverluste bei der Getreideernte)
- aussortierte Fehlprodukte, beschädigte und verunreinigte Produkte (Vgl. [4], S. 87)
- aussortierte Produkte, für die keine Nachfrage existiert, z.B. durch veränderte Modetrends in der Textilindustrie
- Testprodukte und Muster
- Betriebsstoffe (z.B. Schmiermittel, Druckluft, Wärmemittel, Kühlmittel, Energieträger, Reinigungsmittel zur Anlagen Vor- und Nachbereitung) (Vgl. [4], S. 87)

Nicht enthalten sind weitere vermeidbare Verluste, die jedoch im Endprodukt enthalten sind und somit nicht als Materialverluste klassifiziert werden können, wie Materialien die bspw. durch überdimensionierte Produktabmessungen verbraucht werden oder ineffizient ins Produkt integrierte Materialien und Energien. Somit fokussiert die Methode auf das Verhältnis zwischen Inputs, die in den Produktionsprozess fließen und Outputs, die im Endprodukt landen. Demnach sind alle Stoffe, die nicht im Endprodukt verwendet werden, Materialverluste. Nach der Erfassung der Mengen an tatsächlich ins Produkt geflossenen Input-Materialien und der nicht ins Produkt geflossenen Materialverluste, erfolgt die Umrechnung der physischen Mengen mittels monetärer Einheiten in die Kosten der Materialverluste. Dazu werden zunächst die Beschaffungskosten der als Materialverluste klassifizierten Produktionsmittel mit den Kosten für die Entsorgung der Materialverluste summiert. Außerdem werden bestimmte Gemeinkosten anteilig hinzugerechnet. Somit enthalten bspw. die Materialverlustkosten für Schnittreste die Kosten für die Rohstoffbeschaffung, die Energiekosten für den Betrieb der Produktionsanlagen bei der Erzeugung des Verschnitts, die Abfallmanagementkosten, wie auch die Personalkosten für die Erzeugung und die anteiligen Wertverluste der Produktionsanlagen (Abschreibungen) und anderer versteckter Kosten (Vgl. [5], S. 234-235). Wenn eine Maschine in 5% der Betriebszeit Fehlteile produziert, so werden die Kosten für die Wartung der Maschine mit gleichem Anteil den Kosten der Materialverluste hinzugerechnet.

Abb. 1. Exemplarische Darstellung von Materialverlusten über den gesamten Herstellungszyklus eines Produkts

Die Stoffe und Energien werden entlang des gesamten Produktionszyklus über sog. Mengenstellen verfolgt. Eine Mengenstelle kann entweder ein Produktionsprozess oder ein Ort, an dem Bestandsveränderungen durchgeführt werden (z.B. ein Zwischenlager), sein. Damit können sowohl für jede Mengenstelle die Materialverlustkosten einzeln berechnet, als auch eine Summierung der gesamten Materialverlustkosten über den gesamten Produktionsprozess vorgenommen werden.

Sobald die Materialverluste erfasst sind, können mittels der Ergebnisse zielgerichtete Maßnahmen geplant und umgesetzt werden, die dazu führen, dass die kostenintensivsten Verluste vermieden oder zumindest minimiert werden. Mögliche Maßnahmen sind, bspw.:

- Die Nutzung von mathematischen Planungsverfahren zur Verschnittminimierung,
- Die Optimierung der Anlagenumrüstung bzw. von Werkzeugwechseln zur Reduzierung des Betriebsstoffverbrauchs, bspw. durch Total Productive Maintenance (TPM) (Vgl. [4]),
- Die Instandhaltung und Wartung von Maschinen zur Verringerung von Fehlproduktionen,
- Die Optimierung der (Produktions-)Anlagenkonfiguration hinsichtlich eines möglichst effizienten Materialverbrauchs,
- Die Überarbeitung des Produktdesigns im Hinblick auf eine möglichst verlustfreie Herstellbarkeit (Vgl. [4], S. 78),
- Die Verwendung von Produktionsanlagen mit einer hohen Energieeffizienz zur Reduzierung des Energieverbrauchs bei der Produktion von nicht vermeidbaren Materialverlusten,
- Die Nutzung von effektiven Verfahren zur Nachfrageprognose (z.B: Qualitative Prognosen, Zeitreihenprognosen), um möglichst keine Produkte herzustellen, für die keine Nachfrage besteht,
- Die Minimierung von Lager- und Transportvorgängen,
- Die effiziente Nutzung von Verpackungs- und Transportmaterialien,

- Die Verstärkung der Zusammenarbeit mit vor- und nachgelagerten Akteuren in der Produktionskette im Hinblick auf Datenaustausch und die Kosten durch Materialverluste.

Lassen sich die Materialverluste nicht vermeiden oder minimieren, kommt auch noch eine Wiederverwendung in Betracht, bspw. durch das Einschmelzen und Recyceln von Metallspänen. Diese führt ebenfalls zu einer Senkung der Kosten, die aus den Materialverlusten resultieren. Allerdings ist die Wiederverwendung in der Regel signifikant nachteilig gegenüber der Vermeidung von Materialverlusten (Vgl. [4], S. 86-87). Das Recycling führt dazu, dass die Materialien den vorherigen vermeidbaren Arbeits- und Energieaufwand in der Stoffbilanz weiter tragen. Deshalb ist es besser, Produktionsprozesse so anzupassen, dass der Materialverlust und damit der gesamte Materialbedarf des Produktionsprozesses minimiert wird, anstatt Fehlprodukte oder Verschnitt in einem nachgelagerten Prozess erneut zu verwerten (Vgl. [3], S. 24). Das Idealziel der MFCA ist es, die Materialverlustkosten auf null zu senken (Vgl. [2], S. 16). Hierbei gilt die Einschränkung der Wirtschaftlichkeit. Somit sollten die Kosten für die Reduzierungsmaßnahmen die daraus resultierenden Kostenpotentiale nicht übersteigen.

MFCA kann von Beginn an als integrativer Produktionsprozessbestandteil systematisch implementiert werden. Jedoch besteht auch allgemein die Gefahr, dass das Verhältnis zwischen dem Aufwand und dem Nutzen negativ wird, wenn die gesamten Implementierungskosten (also die Kosten der Datenerfassung, Modellierung, Analyse und Maßnahmenumsetzung) die späteren Reduzierungspotentiale übersteigen. Im Regelfall dürfte das vor allem bei weniger materialintensiven Industrien und kleinen Stückzahlen der Fall sein. Da die Anwendung der Methode vor allem bei einem positiven Verhältnis zwischen Aufwand und Nutzen vorteilhaft ist, schlägt die maßgebliche ISO-Norm 14051 zunächst eine Pilotanwendung bei einer Produktionslinie vor, die als besonders ressourcenintensiv gilt (Vgl. [3], S. 30), da dort auch die größten Einsparpotentiale für Materialverluste zu erwarten sind. Nach der erfolgreichen Durchführung, kann mit Hilfe der gewonnenen Erkenntnisse entschieden werden, auf welche anderen Produktionsbereiche sie angewandt wird und ob eine systematische Einbindung in das Produktionsmanagement erfolgt.

4 Ausbreitung und Potential der MFCA-Methode

Obwohl die MFCA-Methode bereits seit den 1980er Jahren existiert und seit 2011 ISO normiert ist, ist die Ausbreitung bisher eher schleppend verlaufen (Vgl. [6], S. 2-6). Recherchen zum Stichwort Materialflusskostenrechnung fördern aktuell hauptsächlich Veröffentlichungen von Wissenschaftlern, Wirtschaftsverbänden (bspw. International Federation of Accountants), Regierungsvertretern und NGOs (UN) zutage. Unternehmensreferenzen gibt es mit Ausnahme einer Studie des japanischen Ministeriums für Wirtschaft, Handel und Industrie (METI) [4] bisher nur in geringer Anzahl. Die Ausbreitung befindet sich momentan in der Diffusionsphase der „Early Adopters" (Initialphase) (Vgl. [6], S. 2-4). Die Methode wird somit noch nicht von einem breiten Anwenderkreis getragen. Ein wesentlicher Grund dafür liegt in der Komplexität der Methode (Vgl. [6], S.5-8). Die Geschwindigkeit, mit der eine Innovation in Unternehmen integriert wird, sinkt mit einer höheren Komplexität und bei einer zunehmenden Voraussetzung von Kenntnissen und Fähigkeiten (ebd.). Für die potentiellen Nutzer ist zunächst nicht erkenntlich, wann der Aufwand für die Einführung der Methode sich durch Kosteneinsparungen amortisieren wird, da vorher nicht erfasst wurde, wieviel Prozent eines bestimmten Inputs nicht im Endprodukt landen, weil sie etwa als Teil von Fehlprodukten von der Qualitätssicherung aussortiert werden. Der Schlüssel zu einer umfassenderen Ausbreitung liegt darin, dass die Methode in den Unternehmen, in denen sie prototypisch angewandt wird, ein positives Kosten-Nutzen-Verhältnis erzeugt und somit einen Wettbewerbsvorteil bewirkt (Vgl. [7], S. 17-29). Die Wettbewerber sind dann aus Wirtschaftlichkeitserwägungen gezwungen, ebenfalls die Methode anzuwenden.

In Japan kennen, begünstigt durch eine Initiative des Ministeriums für Wirtschaft, Handel und Industrie (METI), 80 % der Unternehmen die Methode, es setzen sie aber nur 6,5% ein (Vgl. [8], S. 332). Das lässt darauf schließen, dass die Methode nicht aufgrund von Unkenntnis nicht über die Phase der Early Adopters hinaus gekommen ist, sondern aufgrund von Schwierigkeiten bei der Umsetzung. Das Ministerium hat eine Fallstudie erstellt, in der die Implementierung der Methode von 23 japanischen Unternehmen vorgestellt werden. Davon haben 12 Unternehmen quantitative Angaben zu den Ergebnissen gemacht. Mit einem durchschnittlichen Anteil von 25,9% Materialverlusten an den Gesamtkosten, ist ein bedeutendes Optimierungspotential vorhanden (Vgl. [4], S. 8-84). Von den 23 aufgezeigten Beispielen, finden sich auch fünf Unternehmen, die nicht selbst produzieren, sondern im Dienstleistungsbereich zu verorten sind. Die Beispiele zeigen, dass die Methode zwar generell eher für die Industrieproduktion geeignet ist, aber auch darüber hinaus zu einem Prozessverständnis und zu Effizienzgewinnen führen kann (sog. Awareness). Für die Datenerhebung wird zwar die Expertise von Ingenieuren, Produktionsplanern sowie Buchhaltern gleichermaßen benötigt, die Methode an sich wird aber auch schnell von Nicht-Spezialisten verstanden (Vgl. [6], S. 5). Für die Anwendung der Methode bietet es sich deshalb an, interdisziplinäre Teams zu bilden.

Zwar ist die MFCA prinzipiell eher gut für produzierende Industrien geeignet, jedoch zeigt die Studie des METI [4] in fünf verschiedenen Fallstudien Möglichkeiten auf, die Methode bei Dienstleistungsunternehmen einzusetzen. Dabei werden alle Ressourcen, die bei der Erbringung der Dienstleistung notwendig sind und während des Dienstleistungsprozesses verloren gehen oder unnötig aufgewendet werden, als Materialverluste gewertet, wie bspw. defektes Büroinventar oder nicht verwendete Essensreste in einem Restaurant. Die Methode ist nicht geeignet, die externen Kosten der Materialverluste (z.B. Umweltbelastungen, die aus der Abfallentsorgung resultieren) abzuschätzen und somit die ökologischen Folgen die sich durch die Ineffizienzen ergeben, zu beziffern. Dadurch ist es möglich, dass die Ergebnisse finanzielle Einsparpotentiale aufzeigen, die durch Substitution eines Materialverlustes durch ein anderes Material möglich sind. Diese Substitution kann dann zwar ökonomisch sinnvoll, aber ökologisch nachteilig sein. Dies wäre bspw. der Fall, wenn ein Blei-Säure-Energiespeichersystem aufgrund niedrigerer Kosten durch ein Speichersystem auf Lithium-Eisenphosphat-Basis ersetzt würde, das eine schlechtere Ökobilanz aufweist (Vgl. [9], S. 74 ff.).

Selbst wenn keine signifikanten Einsparungen erzielt werden, bringt die Methode den Vorteil, dass das generelle Verständnis der Produktion erhöht wird und eine grundlegende Transparenz in der Lieferkette erzeugt wird.

5 Herausforderungen bei der Implementierung einer Materialflusskostenrechnung

Vergleichsweise wenige Unternehmen erheben die für eine MFCA erforderlichen Daten systematisch, bezogen auf Materialien, Produkte oder Prozesse, sodass die Methode nicht ad hoc am Schreibtisch angewandt werden kann (Vgl. [8], S. 26-32). Die generelle Berechnung der Kosten, die über die Verkettung von Mengenstellen für Materialverluste anfallen, ist dabei nicht die größte Herausforderung, sondern die Erfassung von bisher unbekannten Mengendaten bzw. das Zusammentragen der Daten aus verschiedenen Quellen. Im Prinzip wäre die Erfassung der relevanten Daten trivial, wenn man zunächst alle Materialien, die als Input in die Produktion fließen, wiegen würde und nach der Produktion die Anteile am Endprodukt erneut wiegen könnte, da die Differenz zwischen beiden Werten die physischen Materialverluste darstellen. Da jedoch während der Produktion Stoffe Verbindungen eingehen und genaue Anteile einzelner Materialien im Endprodukt im Regelfall nicht ohne weiteres zu bestimmen sind, müssen als Materialverluste die Mengen bestimmt werden, die nicht ins Endprodukt geflossen sind. Eine Ausnahme davon bilden Prozesse, bei denen lediglich ein einzelnes Material als Input verwendet wird.

Bei der Datenerfassung wird bisher häufig auf eine maßgeschneiderte Softwareunterstützung verzichtet, stattdessen wird auf allgemeine Tabellenkalkulationsprogramme zurückgegriffen oder mit Zettel und Stift gearbeitet (Vgl. [4], S. 8-84). Prinzipiell kann die Methode auch ohne Softwareunterstützung angewandt werden, da als Ergebnis die Materialkosten in tabellarischer Form dargestellt werden. Die bevorzugte Darstellungsform ist eine nach ISO 14051 normierte Kostenmatrix (siehe Abb. 2).

	Materialkosten	Energiekosten	Systemkosten	Entsorgungskosten	Gesamtkosten
Endprodukt	15.000	1.000	17.000	-	33.000
	38,7%	2,6%	43,8%	-	85%
Materialverluste	3.000	300	2.000	-	5.300
	7,7%	0,8%	5,2%	-	14%
Abfall / Recycling	-	-	-	500	500
	-	-	-		1%
Gesamt	18.000	1.300	19.000	500	38.800
	46,4%	3,4%	49,0%	1,3%	100%

Abb. 2. Materialflusskostenmatrix als Ergebnis der MFCA-Analyse

In der Praxis führt die Methodendurchführung ohne spezialisierte Hilfsmittel aber zu einer Reihe von Umsetzungsschwierigkeiten. Vor allem bei großen Datenmengen und einer dauerhaften Implementierung der Methode wird eine systematische Herangehensweise erforderlich, die vor allem mit Hilfe von Softwareunterstützung erreicht werden kann (Vgl. [4], S. 64.). Die Mengen werden in der Regel über einen vorbestimmten Zeitraum durch Zählung, Wiegung oder andere Messverfahren exemplarisch erfasst. Aus diesen Beispieldaten müssen die Werte für den Betrachtungszeitraum der MFCA-Analyse extrapoliert werden, da eine exaktere Bestimmung häufig nicht möglich ist. Materialverluste können teilweise sehr unregelmäßig auftreten und es fehlt an einfachen Messmethoden, da die Materialverluste nicht automatisch durch die Produktionsanlagen erfasst werden. Durch benutzerfreundliche Softwareanwendungen lässt sich die Komplexität bei der Datenerfassung jedoch senken (Vgl. [10], S. 1-10). In Problemfällen können Hilfestellungen zur Verfügung gestellt werden. So kann eine Anwendung bspw. darauf hinweisen, dass es sinnvoller sein kann mit Referenz- oder Schätzwerten zu arbeiten, anstatt über einen langen Zeitraum exakte Daten zu erfassen und die Methodenanwendung so unrentabel werden zu lassen. Materialien, bei denen geringe finanzielle und ökologische Auswirkungen erwartet werden, können bei der Datenerfassung ausgeklammert werden (Vgl. [2], S. 9). Jedoch nimmt damit auch die Qualität und damit die Aussagekraft der Ergebnisse ab. Es ist also das richtige Maß zu finden zwischen einer kostspieligen und aufwendigen Erhebung exakter Messwerte einerseits und der wirtschaftlichen Implementierung der Methode mittels weniger genauer Messwerte andererseits. Generell lassen sich diese Schwierigkeiten aber verringern, so dass das Potential für eine schnelle Ausbreitung gegeben ist (Vgl. [6], S. 8).

Ein zuvor veröffentlichter Beitrag von Hemke und Wohlgemuth [11] stellt eine Auswahl bestehender Softwareanwendungen zur Durchführung einer MFCA vor. Dazu gehören Umberto Efficiency+ [12] und bw!MFCA [13] der ifu Hamburg GmbH, der Materialflusskostenrechner des VDI Zentrum Ressourceneffizienz [14] sowie Excel-Kalkulatoren, die jedoch in japanischer Sprache verfasst sind [15]. Die Software-Lösungen sind jedoch noch nicht auf dem Entwicklungsstand, der die genannten Herausforderungen für Unternehmen senkt und eine systematische Umsetzung begünstigt. Einige der Kritikpunkte sind:

- Die fehlende Unterstützung bei der Datenerfassung,
- geringfügige Unterstützung zum Verständnis der Methode,
- unzureichende grafische Auswertungsmöglichkeiten,
- die Vernachlässigung der Vorgaben der ISO 14051,
- fehlende Verknüpfung mit anderen Datenbeständen mittels Schnittstellen,
- ein, über die MFCA hinaus gehender, Funktionsumfang und damit verbundene unnötige Anschaffungskosten, sowie
- eine fehlende Unterstützung verschiedener Einsatzszenarien (z.B. Pilotstudie).

Gleichwohl treffen diese Kritikpunkte nicht generell auf alle genannten Anwendungen zu.

6 Softwareanforderungen zur systematischen Umsetzung der MFCA-Methode

Eine Reihe von Anforderungen, deren Umsetzung mittels Software zu einer Vereinfachung der Datenerfassung und Modellbildung förderlich ist, wird im folgenden Abschnitt zusammenfassend beschrieben. Zur Ermittlung wurde von den Autoren eine Usability-Studie durchgeführt, bei der sechs Testpersonen typische Aufgaben im Rahmen einer Materialflusskostenrechnung erledigt haben. Die Testpersonen setzten sich aus Firmenvertretern und Masterstudenten der Umweltinformatik zusammen, bei denen eine Affinität zu dem Thema gegeben ist. Ziel war es, Hinweise zu erarbeiten, wie die Methode effektiv, effizient und zufriedenstellend umgesetzt werden kann. Die sechs teilnehmenden Testpersonen verfügten bereits Vorkenntnisse aus dem Bereich der Umweltinformatik. Der Ablauf und weitere Ergebnisse werden von Hemke und Wohlgemuth [11] vorgestellt. Weitere Anforderungen wurden mittels einer Anforderungsanalyse ermittelt.

6.1 Datenerfassung

Zunächst sollte die noch zu entwickelnde Anwendung den Spagat bewältigen, den die Empfehlung der ISO 14051 nahelegt, nämlich die Methode zunächst testweise in einer kleinen Pilotphase einzusetzen und mit den gewonnenen Erkenntnissen diese dann in weiteren Unternehmensteilen bis hin zum gesamten Unternehmen zu implementieren. Dadurch sollte sie es ermöglichen, sowohl kleine Ketten von Mengenstellen übersichtlich zu modellieren, als auch große Eingabemasken für die Darstellung von globalen Produktionsketten zur Verfügung stellen. Sobald das Modell größere Ketten von Mengenstellen abbildet, muss eine Skalierung der Benutzungsoberfläche dafür sorgen, dass die Übersichtlichkeit nicht verloren geht (bspw. mittels Responsive Design). Sind die Mengenstellen für das Untersuchungsobjekt vollständig erfasst, müssen zu diesen Mengenstellen jeweils die genutzten Materialtypen der Input-Stoffe, deren Mengen und Kosten gesammelt werden. Bei der Angabe von Materialtypen können Materialdatenbanken eine Hilfestellung geben (wie z.B. MatWeb [16]). Anschließend muss der Anteil der Input-Materialien bestimmt werden, der nicht ins

Endprodukt fließt und damit als Materialverlust zu werten ist. Die Erfassung dieser Mengen ist in der Regel der Knackpunkt bei der MFCA-Analyse, da Unternehmen vorher noch nicht über alle notwendigen Daten und Erfahrungen verfügen und es mitunter zu Schwierigkeiten oder Fragen kommt, etwa:

- Wieviel Kilogramm wiegt ein 7 m³ Container in dem sich Verschnitt befindet?
- Ist die aktuell gemessene Menge an Fehlteilen repräsentativ für den Betrachtungszeitraum?
- Wieviel Strom hat die Anlage zur Herstellung von einem Produkt verbraucht?
- Welche Menge von Hilfsstoff X wird exakt für die Herstellung eines Produkts benötigt?
- Wie werden Materialmengen, die in unterschiedlichen Einheiten gemessen werden, auf eine gemeinsame Einheit umgerechnet (z.B. von Liter in Kilogramm)?

Der Aufwand zur Datenerfassung wird sich nicht gänzlich durch eine Software beheben lassen, aber er kann gesenkt werden, in dem die Anwendung eine möglichst große Bandbreite an Hilfestellungen für diesen sensiblen Punkt in der Materialflussanalyse bereitstellt. Hierzu gehören etwa Durchschnittswerte für die Umrechnung von Volumen zu Masse, die softwareseitige Unterstützung zur Datenintegration von mobilen Messsensoren (digitale Wagen, Smart Meter etc.) und Datenschnittstellen zu bestehenden Datenbanken und ERP-Systemen. Die Ergebnisse bei der Nutzung von Durchschnittswerten und Näherungswerten entsprechen zwar nicht immer exakt der Realität, sie sind jedoch der Alternative, keine Daten zu erheben und damit auch keine Materialflusskostenrechnung durchzuführen, überlegen. Auch begründete Schätzwerte können ein wichtiger Baustein sein, um sonst nicht überbrückbare Lücken in der Datenerfassung auszugleichen (Vgl. [17], S. 20-31). Die Materialien müssen in unterschiedlichen Einheiten angegeben werden können, so dass eine interne Umrechnung stattfindet. Gleiches gilt für die Währung, in denen die Materialien bezogen wurden.

Eine besondere Hürde besteht außerdem im Ausgleich der Bilanz zwischen Input-Mengen und Output-Mengen einer Mengenstelle. So muss die Gesamtmasse der Materialien, die in der Produktion eingesetzt werden, genau so groß sein, wie die Gesamtmasse des Endprodukts und der Materialverluste. In der Praxis kommt es hierbei häufig zu Problemen, da es zu Rundungsfehlern kommen kann oder Materialien, von denen nur geringe Mengen genutzt werden, vergessen werden. Für diese Hürde ist eine Hilfestellung zu konzipieren, etwa indem die Anwendung selbst ermittelt, in welcher Mengenstelle es zum Konflikt gekommen ist, wie groß die Abweichung ist und wie sie zu beheben ist.

Zusätzlich zu den direkten Kosten, die die Beschaffung der Input-Materialien verursacht hat, müssen allgemeine Systemkosten bzw. Prozesskosten erfassbar sein. Zu diesen zählen etwa Energiekosten, die Abfallmanagementkosten und andere Kosten (etwa Abschreibungen auf Anlagen, Personalkosten, Instandhaltungskosten etc.). Diese

Systemkosten, werden anteilig den Materialverlusten hinzugerechnet, so dass die Kosten der Materialverluste bspw. zuletzt auch den Anteil der Personalkosten enthalten, der für die Produktion von Fehlprodukten aufgewendet wurde. Vorteilhaft ist an diesem Punkt, eine Eingabemöglichkeit sowohl von absoluten, als auch von prozentualen Anteilen (Vgl. [14]).

Projektteams, die mit der Umsetzung betraut werden, sollten einen Querschnitt des Unternehmens abbilden, bestehend aus der Unternehmensleitung, der Buchhaltung, der Produktionsplanung, der Umweltabteilung und Produktionsmitarbeitern, die mit dem genauen Ablauf der Produktion vertraut sind und die Änderungen im Produktionsprozess umsetzen müssen (Vgl. [2], S. 12). Insofern muss die Anwendung für Benutzer mit unterschiedlichen Sichtweisen und Kenntnisständen gleichermaßen konzipiert sein.

6.2 Datenauswertung

Die Darstellung der Ergebnisse erfolgt generell in Form der MFCA-Kostenmatrix nach ISO 14051 (siehe Abb. 2), diese sollte daher auch den Kern der Datenauswertung darstellen. Zusätzlich sind jedoch weitere Auswertungsfunktionen denkbar, die die Aussagekraft der Ergebnisse steigern können bzw. bei großen Analysen bestimmte Ausschnitte leichter ersichtlich machen. Eine Aufschlüsselung der Materialverlustkosten je Mengenstelle kann dabei eben so viel beitragen wie eine Nachvollziehbarkeit der Verluste eines spezifischen Materials über mehrere Mengenstellen hinweg. Zur grafischen Veranschaulichung des Mengenstellenmodells bietet sich ein Materialflussmodell an, das die Mengenstellen und deren Zusammenhänge darstellt. Die Kostenströme können durch Sankey-Diagramme visualisiert werden, die die Kosten durch größenproportionale Pfeilen darstellen. Die Ergebnisse einer MFCA sind wichtige Kommunikationsmittel bei der Managementunterstützung wie auch in der Außendarstellung (Vgl. [3], S. 16). Demzufolge bieten Möglichkeiten zur grafischen Aufbereitung und Verdichtung der Ergebnisse (z.B. mittels Kennzahlen) einen wichtigen Vorteil bei der Nutzung der Methode. Die Auswertungsmöglichkeiten müssen dem Ziel der Methode, nämlich unnötige Materialverluste aufzudecken, angepasst werden.

6.3 Weitere Anforderungen

Neben der Darstellung der Ergebnisse, sollte die Anwendung Hilfestellungen zur Ergebnisinterpretation geben und den Benutzern somit einen Einstieg in die Reduzierung der Materialverluste geben (z.B. in Form von Best Practices oder Vergleichsstudien). Das größte Potential zur Reduzierung bringen in der Regel die Mengenstellen mit den höchsten Kosten durch Materialverluste. Hierbei sollten dem Benutzer aber Hilfestellungen zur Ergebnisinterpretation und zum Verständnis der Methode bereitgestellt werden. Für Materialverluste, die sich nicht beseitigen lassen, sollte die Anwendung Möglichkeiten zum Recycling vorschlagen. So könnte bspw. eine Vermittlung von mehreren Unternehmen hin zu einer Industriesymbiose erfolgen, so dass die Abwärme eines Unternehmens von einem anderen Unternehmen als Prozessinput verwendet wird (Vgl. [18]). Um eine parallele Datenhaltung zu vermeiden und die erhobenen Daten auch in anderen Kontexten nutzen zu können (bspw. Carbon Footprinting), sollte die Anwendung eine Funktionalität zum Daten Import und Export enthalten. Weitere administrative Anforderungen, wie das Festlegen von Rollen und Zuständigkeiten, das Festlegen der Systemgrenzen und die Erfassung von Erfahrungswerten (Lessons Learned) werden von der Asian Productivity Organization [2], S. 17-25 genannt. Sinnvoll ist zu diesem Zweck eine Sammlung von Best Practices in Form eines Repository.

7 Zusammenfassung und Ausblick

Die aktuelle Situation bei der Ausbreitung der Materialflusskostenrechnung kann als ambivalent eingeschätzt werden. Einerseits bietet die Methode ein großes Potential, die unternehmerische Effizienz zu erhöhen und damit die Kosten bei gleichbleibendem Sinken der negativen Umweltauswirkungen zu reduzieren. Eine ausführliche Pilotstudie in Japan, die auf einer Initiative des METI basiert, zeigt die großen Einsparpotentiale auf (Vgl. [19], S.401-407). Abgesehen davon bleibt die Ausbreitung bisher aber hinter den Erwartungen zurück. Als wichtiger Grund dafür ist die Notwendigkeit der Erfassung von Daten, die bisher in der erforderlichen Form selten vorliegen. Der Aufwand lässt sich schwer gegen einen vermuteten Nutzen abwägen, zumal eine Befragung ergeben hat, dass Unternehmen häufig der trügerischen Überzeugung sind, dass bei ihnen keine unnötige Verschwendung stattfindet (Vgl. [4], S. 91), Fallstudien jedoch das Gegenteil zeigen (Vgl. [20], S. 7-8).

Da bei der Datenerhebung, je nach Komplexität der Produktion, sehr große Datenmengen zu managen sind, ist ein systematischer Ansatz bei der Erfassung sowie der MFCA Berechnung notwendig (Vgl. [4], S. 64), der durch eine zielgerichtete Softwareunterstützung sicher gestellt werden kann. Die beschriebenen Schwierigkeiten bei der Datenerfassung sollten im Fokus der Anwendung stehen, da hier die größten Potentiale bestehen, die Methodenanwendung zu vereinfachen. Der vorliegende Beitrag hat in einer ersten Studie eine Vielzahl von Softwareanforderungen ermittelt, die jedoch noch keine abschließende Auflistung darstellen. Die bestehenden Anwendungen decken schon Teile der Anforderungen ab, liefern aber keine zufriedenstellende Gesamtlösung. Sie sind jeweils nur auf bestimmte Anwendungsfälle fokussiert. Außerdem bieten die zur Verfügung stehenden Softwarelösungen an neuralgischen Punkten, wie der Datenerfassung, wichtige Hilfestellungen nicht an, so dass die Benutzer auf externe Hilfsmittel angewiesen sind, oder unnötigen Mehraufwand in Kauf nehmen müssen. Eine Softwarelösung zur Durchführung der MFCA, die an den Bedürfnissen der Benutzer ausgerichtet ist und damit über eine gute Usability verfügt, kann die Kosten der Methodenumsetzung deutlich reduzieren und damit die Vorteile der Methode besser nutzbar machen.

In folgenden Studien, ist geplant, mittels Interviews und Marktforschung mit Beratern, die bereits Erfahrung bei der Einführung der Methodik und der ISO-Norm besitzen, da hier das größte Potential besteht, die Anforderungen an Hilfestellungen bei der Implementierung der Methodik zu ermitteln und Einblick in die Akzeptanz im Managementbereich zu erlangen (Vgl. [6], S. 8). Welche Treiber und Hindernisse sind bei der Implementierung festzustellen? Wie können die Treiber der Methode durch eine benutzerfreundliche Anwendung zur Geltung gebracht und die Hindernisse abgeschwächt werden? Zur Bearbeitung dieser Fragen werden zukünftig weitere Untersuchungen folgen. Angedacht sind Interviews und Umfragen unter Unternehmensvertretern, die in die Förderung und Anwendung der MFCA-Methode involviert sind. Diese Untersuchungen werden vor allem dann verwertbare Informationen liefern, wenn die Methode in Zukunft weitere Ausbreitung und Bekanntheit erlangt.

Literaturverzeichnis

1. Nakajima and Michiyasu, "On the Differences between Material Flow Cost Accounting and Traditional Cost Accounting - In Reply to the Questions and Misunderstandings on Material Flow Cost Accounting," in Kansai University Review of Business and Commerce, Suita, Kansai University, 2004. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
2. Asian Productivity Organization, "Manual on Material Flow Cost Accounting ISO 14051," Tokyo, 2014.
3. DIN Deutsches Institut für Normung, DIN EN ISO 14051: Environmental Management – Material Flow Cost Accounting – General Framework (ISO 14051:2011), Berlin, 2011.
4. The Ministry of Economy, Trade and Industry of Japan, "Material Flow Cost Accounting MFCA Case Examples," Japan, 2010.
5. A. Schmidt, B. Hache, F. Herold and U. Götze, "Material Flow Cost Accounting with Umberto®," R. Neugebauer, U. Götze and W. Drossel, Eds., Auerbach, Verlag wissenschaftliche Scripten, pp. 231–248.
6. K. L. Christ and R. L. Burritt, "ISO 14051: A new era for MFCA implementation and research," in Revista de Contabilidad – Spanish Accounting Review 19, 2014, pp. 1-9.
7. P. Gorecki and P. Pautsch, Lean Management, München: Carl Hanser Verlag, 2018.
8. K. Kokubu and E. Nashioka, "Environmental Management Accounting Practices in Japan," in Implementing Environmental Management Accounting: Status and Challenges, Dordrecht, Springer, 2005, pp. 321-343.
9. A. R. Köhler, Y. Baron, W. Bulach, C. Heinemann, M. Vogel, S. Behrendt, M. Degel, N. Krauß and M. Buchert, "Ökologische und ökonomische Bewertung des Ressourcenaufwands - Stationäre Energiespeichersysteme," VDI Zentrum Ressourceneffizienz GmbH, Berlin, April 2018.
10. M. Richter and M. Flückiger, "Usability Engineering kompakt 3. Auflage," Berlin Heidelberg, Springer Vieweg, 2013.
11. F. Hemke and V. Wohlgemuth, "Analyse aktueller Softwareanwendungen zur Durchführung einer Materialflusskostenrechnung (MFCA)," in Proceedings des 6. Workshop Umweltinformatik zwischen Nachhaltigkeit und Wandel, 2018.
12. ifu Hamburg GmbH, "Umberto Efficiency+," https://www.ifu.com/umberto/effizienz-software/. (Accessed 08 2018).
13. ifu Hamburg GmbH, "bw!MFCA," https://www.ifu.com/umberto/bwmfca/. (Accessed 08 2018).

14. VDI Zentrum Ressourceneffizienz, "Materialflusskostenrechner," 2014. https://kostenrechner.ressource-deutschland.de/#/materialverluste. (Accessed 08 2018).
15. JMA Consultants Inc, "MFCA Thinking," 2010. http://www.jmac.co.jp/mfca/thinking/07.php. (Accessed 08 2018).
16. MatWeb, "Material Property Data," 2018. http://www.matweb.com/. (Accessed 08 2018).
17. Bundesumweltministerium / Umweltbundesamt, "Leitfaden Betriebliches Umweltkostenmanagement," Berlin, 2003.
18. A. Lütje, A. Möller and V. Wohlgemuth, "A Preliminary Concept for an IT-Supported Industrial Symbiosis (IS) Tool Using Extended Material Flow Cost Accounting (MFCA) - Impulses for Environmental Management Information Systems (EMIS)," in Advances and New Trends in Environmental Informatics, Garching, Springer Nature Switzerland AG, 2018.
19. Y. Onishi, K. Kokubu and M. Nakajima, "Implementing Material Flow Cost Accounting in a Pharmaceutical Company," in Environmental Management Accounting for Cleaner Production, S. Schaltegger, M. Bennett, R. L. Burritt and C. M. Jasch, Eds., Springer, 2008, pp. 295-409.
20. K. Kokubu and H. Kitada, Conflicts And Solutions Between Material Flow Cost Accounting And Conventional Managemnt Thinking [sic], Sydney, 2010.
21. M. Schmidt and M. Nakajima, "Material Flow Cost Accounting as an Approach to Improve Resource Efficiency in Manufacturing Companies," resources, no. 2, pp. 358-369, 2013.
22. S. Geib, "VDI-Z Integrierte Produktion," 03 2018. https://www.vdi-z.de/2018/Ausgabe-03/Sonderteil-Materialfluss/Mehr-Transparenz-beim-Materialfluss. (Accessed 08 2018).

Computing Incentives for User-Based Relocation in Carsharing

Bernd Herrenkind[1], Alfred Benedikt Brendel[1], Sascha Lichtenberg[1], and Lutz M. Kolbe[1]

[1] University of Goettingen, Chair of Information Management, Goettingen, Germany
bernd.herrenkind@stud.uni-goettingen.de,abrendel@uni-goettingen.de, sascha.lichtenberg@stud.uni-goettingen.de, lkolbe@uni-goettingen.de

Abstract. Carsharing offers an environmentally friendly alternative to private car ownership. However, carsharing providers face the challenging task of matching shifting vehicle supply with fluctuating customer demand to prevent related operational inefficiencies and ensure customer satisfaction. To date, researchers have improved existing relocation strategies and developed new concepts with the use of information technology tools. Still, current literature lacks research on optimization and implementation of user-based relocation solutions. The most urgent need currently lies in the development of algorithms to compute and implement effective incentives for user-based relocation.
We address these needs by utilizing a design science research approach to develop an automated machine learning-based incentive computation solution for incentivizing user-based relocation. We use a survey of 274 participants resulting in 1370 individual data points to train an incentive computation model, which is then applied within a small-scale field test. Results suggest that the algorithm computes appropriate incentives.

Keywords: Design Science Research, Carsharing, User-based Relocation, Incentive Computation, Green IS

1 Introduction

The demand for new sustainable mobility options as substitutes to private vehicles has increased in recent years. However, to be considered a valid alternative, new mobility options should not only be sustainable but also be flexible enough to meet dynamic customer and environmental needs. In this context, the use of alternate mobility services can be a possible solution, with carsharing as a prime example [1–3].

Current research in the area of carsharing views balancing vehicle supply and demand as one of the key enablers for this to happen [4–6]. Carsharing providers face the challenge of a shifting vehicle supply and demand, leading to insufficient vehicle supply at some locations. Consequently, some customers may request a vehicle, but may have their rental request rejected [7–9]. This situation reduces revenue and leads customers to view carsharing as inflexible and inferior to a privately owned vehicle [5].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Hence, there is a need for carsharing providers to take countermeasures to sustain a sufficient vehicle distribution. The common counter-measure is to relocate vehicles in advance by operator-based relocation [6, 10], a process in which employees move vehicles between stations or areas to comply with local demands [3, 5, 6]. However, this procedure has high costs (e.g. personnel and fuel) associated with it, while lowering these operating costs remains a key success factor for carsharing [3, 9, 11].

Therefore, to provide an alternative researchers have developed a new concept called user-based relocation. The idea of user-based relocation is to motivate users to return currently rented vehicles to locations with high vehicle-demand [4, 12]. This relocation approach is deemed to be less expensive [3, 12], but is yet to be implemented within a real-world commercial carsharing system. Inducing user-motivation to perform relocation using the right mechanisms, e.g. offering monetary incentives, is a major obstacle in this process, as methods to compute sufficient incentives are still lacking [3, 4, 13]. Current carsharing research often uses a simplified function to compute incentive costs for their scenarios [4, 12] without real-world application. However, the incentive must be derived from various factors, such as weather, distance or time. To close this research gap, we aim in this paper to answer the following question:

RQ: How can a cost-efficient incentive be computed for user-based relocation?

2 Related Research

Carsharing systems can be distinguished into three forms [6]:
(1) **Station-based two-way carsharing:** Vehicles can be rented by customers from any station and the vehicles have to be returned to the same station.
(2) **Station-based one-way carsharing:** Vehicles can be rented from any station but can also be returned at any available station.
(3) **Free-floating carsharing:** Unlike station-based carsharing, free-floating carsharing allows cars to be rented and returned anywhere within the operation area of the carsharing provider.

The flexibility of station-based, one-way, and free-floating carsharing possesses a drawback in the form of the vehicle relocation problem – vehicle distribution is altered by rentals, leading to an under-supply of vehicles at some stations or in some areas. Hence, carsharing providers have to manage vehicle distribution by vehicle relocation to achieve a more balanced system [14].

Currently, vehicle supply and demand management measures can be distinguished into three forms:
(1) **Operator-based relocation:** Staff members rearrange the vehicles by driving, towing or ride-sharing them to the desired location [15].
(2) **User-based relocation:** Selected users are motivated by the carsharing provider during a rental to return their currently rented vehicle at a station with a high vehicle demand, when their initial destination has sufficient vehicle supply [12].
(3) **Prevention-based:** Through territorial pricing strategies and other price incentives, attempts are being made to proactively encourage users to use certain vehicles and park

them at selected locations in order to maintain a balanced supply-demand ratio with regard to vehicle locations [16, 17].

We conducted a literature review [18] to evaluate the current status quo of user-based vehicle relocation research and identify relevant research gaps. For this, we conducted a tailored keyword search to ensure that the results fit the carsharing domain and user-based concept within the following databases: ScienceDirect, EbscoHost, JSTOR, and AIS electronic library. Variations of the following search query were used:

((“car sharing” OR “carsharing”) AND (“relocation” OR “rebalancing” OR “allocation”) AND (“user-based” OR “user based” OR “userbased”))

After filtering, the search resulted in 13 separate publications. Next, we did a forward and backward search, identifying 10 additional publications. All together, we gathered 23 publications on the topic of user-based relocation (see Table 1). In the following, we will present the incentive computation methods for user-based relocation that we found through our research.

Brendel et al. [12] presented a function of time for relocation to compute the incentive. They derived the function from a survey of 26 participants. However, the sample is relatively smaller and lacks an evaluation. Furthermore, only using time for relocation simplifies the problem significantly.

Wagner et al. [4] developed an incentive concept based on the idle time reduction a relocation grants. The user is offered free-minutes in relation to the idle time reduction. However, they do not specify a method to determine the amount of time required to motivate users to relocate.

Angelopoulos et al. [17] call their approach user-based but it does not fit the common definition of user-based relocation. Their approach refers to a dynamic allocation of vehicles to users, based on incentive systems that use reservations to manage mismatches between supply and demand, to create price incentives for rewarding users when they agree to fetch their vehicle from an oversupplied station and/or hand it over to an undersupplied station. On that account, it is a preventive vehicle supply and demand management approach. The other articles found are mainly concerned with providing technical (e.g. algorithmic) support for future implementations of user-based relocations.

Table 1. Literature Overview - Carsharing Vehicle Relocation Research

Article	*Incentive Computation Method*	*Free-Floating*	*Station-Based*
Angelopoulos et al. [17]	**(X)**		**X**
Bannerjee et al. [19]			**X**
Barth and Todd [15]			**X**
Barth et al. [20]			**X**
Bianchessi et al. [21]			**X**
Brandstätter et al. [22]			**X**
Brendel et al. 2016 [12]	**X**		**X**
Cepolina and Farina [23]			**X**
Cepolina and Farina [24]		**X**	**X**
Chow and Yu [25]			**X**

Clemente et al. [26]			X	
Clemente et al. [3]				X
Di Febbraro et al. [27]				X
Gavalas et al. [28]			X	X
Herrmann et al. [29]			X	
Jorge et al. [30]				X
Jorge and Correia [9]J				X
Laarabi et al. [31]				X
Schulte and Voß [13]			X	
Wagner et al. [4]		**(X)**	X	
Wang et al. [32]			X	
Waserhole et al. [33]				X
Wang et al. [34]			X	
	∑ ***23***	***3***	***8***	***17***

X ≙ covers topic completely; (X) ≙ addresses topic partially

For instance, Wagner et al. [4] develop a support system for user-based relocation within free-floating carsharing, which detects imbalances in vehicle distribution based on idle time patterns. Discounted alternative stations are simulated. Bianchessi et al [21] present the “Feedback Dynamic Pricing” approach, which is based on modelling a vehicle sharing system as a dynamic system, aiming to control fleet balancing by varying the price of the service in real time. For this purpose, a simulator was developed. Both articles stand exemplary for the current focus of user-based relocation research. Novel and sophisticated algorithms for detecting and counteracting vehicle imbalances are developed, while necessary incentives are often a comparative side note and not investigated as profoundly.

In summary, current research has not engaged with the topic of incentive prediction from a practical standpoint. The approach of Brendel et al. [12] comes closest to a practical application. However, their study focuses on estimating the potential costs of user-based relocation and is not intended for practical application. Thus, it is visible that developing a practice-ready incentive computation method for user-based relocation is still an evident research gap.

3 Research Approach

We followed a combination of the frameworks by Hevner et al. [35] and Hevner [36]. The DSR setting and its interrelated cycles are depicted in Figure 1. The relevance cycle connects the design activities of the design cycle with the artifact's intended environment. By this, it enables researchers to assemble real-world requirements to describe and later solve the subsequent real-world problems. Furthermore, artifacts are introduced to the environment as part of a relevance cycle. The rigor cycle relates the design activities to the existing body of knowledge. Thus, existing knowledge can be integrated into design activities and research results can later extend the knowledge base. The central part of a DSR process is the design cycle. It represents the design and evaluation activities of the researcher.

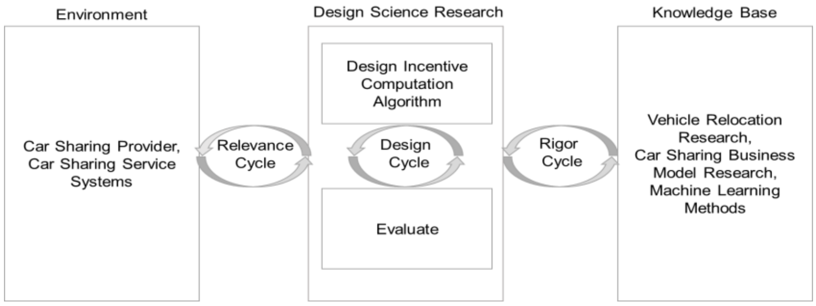


Figure 1. Design Science Research Setting (following [32])

Additionally, during the research process, we also applied a theorizing process similar to heuristic theorizing [37] as described by Brendel, Brennecke, et al. [16] for the development of a design theory. The development of a six component design theory [38] is achieved by implementing the following steps after each research activity:
(1) Review the current design theory and evaluate the need for refinement of each of the six components.
(2) If refinement is needed, iteratively add new components and/or adjust existing components until all new knowledge is incorporated.

For this, we abstracted and de-abstracted the requirements, the development process and the artifact to obtain context specific or meta-components [37].

The DSR process started with a relevance cycle (see Table 2), revealing a lack of implemented user-based relocation system and methods to compute incentives for user-based relocation (see the Literature Review section). To validate our findings, we discussed these results with two carsharing providers. During both the interviews with the managing directors, we presented and discussed the concept of user-based relocation and the related research gap. For this, we prepared the following open questions:
(1) What general limitations and restrictions come to your mind when you think about incentives and user-based relocation in general?
(2) What are the requirements to be considered for the development of an algorithm that can compute appropriate incentives for user-based relocation?

Our findings were confirmed, resulting in the identification of a practice relevant problem manifested as a lack thereof incentive prediction methods for user-based relocation. Furthermore, we gathered requirements for the algorithm design process.

As a second step, we performed a rigor cycle to draw from existing vehicle relocation publications. While seeking a starting point for the design process, we identified machine learning as the basis for our development. This decision was made based on the potential of machine learning [39] and its already successful application in the contexts of carsharing and vehicle relocation [6, 13]. Machine learning can identify patterns within given datasets and predict values based on them. Therefore, it can be considered a potential approach for developing an incentive computation algorithm.

Table 2. Overview of Performed Relevance, Design and Rigor Cycles

	Relevance Cycle	*Rigor Cycle*	*Design Cycle*
Inputs	• Publication databases	• Research database	• Machine learning literature
Methods	• Literature review • Expert Interviews	• Literature review	• Survey • Field study
Steps	• Gather publications from literature reviews • Analyze publications • Discuss findings with experts • Gather requirements	• Analyze publications • Identify input knowledge for design process	• Identify requirements • Design machine learning algorithm • Gather dataset • Train algorithm • Test algorithm in field study
Results	• Research database • Identification of a lack of incentive prediction methods for user-based relocation	• Identification of machine learning methods as a possible solution	• Machine learning algorithm for incentive computation

In the following design cycle, we developed an algorithm based on machine learning and evaluated it by training and testing it with a dataset. As datasets for incentives in the context of user-based relocation are currently not available and cannot be referenced, we had to gather a dataset via survey.

The survey included 1370 data points gathered from 274 survey participants (their characteristics are depicted in Table 3). They were questioned on the street, following a structured questionnaire. The questions were designed to mimic the setting of user-based relocation. In user-based relocation, users walk the extra distance from their new vehicle destination to their previously intended one. Hence, they had to evaluate their possible incentive by factoring distance, weather, weekday and time of the day. These factors were derived as an initial explorative step with the partner company and based on accessibility. Other potentially helpful data sources (such as user data from customers) were inaccessible to us.

Table 3. Characteristics of Survey Participants

Characteristics	*Property*				
Age	< 25 46.31%	26-35 33.22%	36-45 4.36%	46-55 9.73%	> 56 6.38%
Education	No Education 3.42%	(High) school 36.99%	Some college or profession 10.62%	College with degree 45.21%	Higher education or PhD 3.77%
Gender	Male: 58.19%			Female: 41.81%	

The partner carsharing company has mainly younger customer base and is located within a University student majority city. For this reason, it is conclusive that the sample is also relatively young, reflecting the actual average age of customer base. Furthermore, we would like to note that we had not specifically targeted nor filtered our questionnaire for younger participants.

Each participant was asked if they would accept a certain incentive to walk a given distance (between 800m and 1,500m) on a specific weekday. The incentive was presented to the participants in a sale by bidding format, starting from 0.50€ and going up to 6€ [12] in 0.50€ steps. Additionally, the participants were distributed with the weather conditions for the walk. The weather was characterized by: (1) temperature, and (2) weather condition (rainy, windy, snowing, cloudy etc.). In total, each participant was surveyed for five different combinations. The first one was the current time and place's weather conditions. In order to ensure a variety of different real-world weather conditions within our dataset, we conducted the survey in two time-windows: one week in April and one week in October 2016. For the other four, each parameter combination was generated randomly, while any sort of illogical combination (e.g. snowy and 30°C) was discarded. Further, each question was asked by corresponding researcher while the participants did not actively fill out the question themselves (preventing wrong inputs within the questionnaire).

We chose to apply a threshold for the maximum incentive and also to utilize the bidding mechanism based on a preceding pre-test. In a small pre-test (of 40 participants), we discovered that participants answered with either very high or low values, when not given a context in the form of a limit. For example, some participants stated that they would like to get an incentive of 50€ or more to walk 800m in good weather conditions. Hence, we limited the participants by putting a cap of 6€ on the incentive (assuming that users demanding an incentive of over 6€ would not participate in user-based relocation) and also by using a bidding style questionnaire. Furthermore, each participant had the option of answering with the statement that they were unwilling to walk the distance under the given circumstances for under 6€. This framework greatly improved the consistency of the gathered data.

The trained algorithm was then evaluated. Normally, machine learning algorithms are evaluated using error metrics. However, as the aim of the developed algorithm is not to predict a value, but rather to iteratively learn to compute incentives, we had to use a different approach. Firstly, the algorithm was trained using the survey data. Secondly, it was evaluated in a small field study. In the field study, 20 carsharing customers were asked to relocate their vehicle during an ongoing rental.

4 Results

4.1 Incentive Computation Algorithm

Building on current user-based vehicle relocation publications [4, 12] and the discussion with carsharing providers, it is evident that despite being an interesting option, implementation of user-based relocation is difficult in the fact that it has to be set within existing infrastructure. Nonetheless, focusing on incentives as key enablers of the concept, we identified the following requirements:

R1 Self-learning: The algorithm must be self-learning to avoid manual adjustments of the computation algorithm.

R2 Context sensitive: The algorithm must consider external factors (like weather), which influence the decision making of customers.

R3 Real time: The incentives must be computed in real-time to achieve a dynamically working user-based relocation system.

R4 Cost-efficient: The algorithm must avoid overpaying customers and paying more than the cost of operator-based relocation.

R5 Simple to implement: The algorithm must be easy to implement to ensure adaptation by other carsharing providers.

R6 Adaptive: The algorithm must be adaptable to different carsharing systems.

R1 addresses the need for an incentive computation algorithm to be self-learning. The customer base of a carsharing system changes over time and incentives must be context sensitive to address price elasticity. Additionally, short-term influences (such as weather) can change the needed incentive and therefore, should be considered during computation (R2). Similarly, an incentive computation algorithm must provide incentives in real-time (R3). User-based relocations must be communicated quickly to potential relocating users to ensure the immediate execution of the relocation. Otherwise, potential users could arrive at their destination to find no cars available or the vehicle distribution could change, leading to relocations that cannot be performed or are unnecessary or even counterproductive. R4 captures the main objective of relocation, as well as the perspective of the carsharing provider. Vehicle relocations must be cost efficient, meaning that the incentives for user-based relocation should be as low as possible, avoiding any overpayment. This allows full leverage of the cost-saving potential of user-based relocation [3, 12]. Due to the fact that carsharing providers are mainly concerned with the "physical" part of carsharing, the implementation of complex algorithms is outside the range of their core-competencies. To ensure a wide adoption of an incentive computation algorithm, it must be designed with an easy implementation in mind (R5), which can maintain an easy and efficient cohesion with the existing infrastructure. Lastly, R6 captures the primary intentions of DSR, which is to develop artificial solutions and to understand solution design on an abstract level, e.g. as design principles [37, 38, 40]. Therefore, the incentive computation algorithm must be applicable in different carsharing systems.

Building on the requirements and a search for relevant knowledge, we decided to use machine learning and apply a regression tree. Machine learning algorithms can find patterns within a given dataset (R1, R6). A regression tree is a well understood and commonly applied algorithm, which has been implemented in one of the most common machine learning programming libraries, e.g. scikit-learn [39], making it simple to implement (R5). It can incorporate different context inputs provided in the dataset (R2, R6) to quickly (R3) compute a numeric output. The cost-efficiency of the computed incentive depends on the provided input data set. If the algorithm can locate robust patterns within the given dataset, the computed incentive will be as cost-efficient as possible (R4). As we used Scikit-learn [39] for our implementation (R5), the Classification and Regression Tree (CART) was used.

In the following design cycle, the gathered requirements and the identified regression tree algorithm were combined to develop an incentive computation algorithm (see Figure 2).

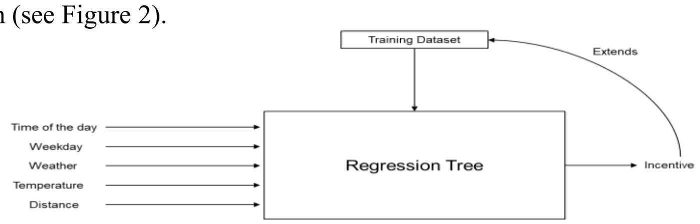


Figure 2. Incentive Computation Algorithm

In the center of the incentive computation algorithm, the regression tree machine learning model is displayed. It is trained via a training dataset (for further information see [41]) and computes incentives based on the time of the day, day of the week, weather, temperature and relocation distance. As a result of the algorithm requiring to adapt to the potentially shifting price sensitivity of the customer base, it likewise requires itself to be self-learning (R1). Hence, we designed the following process: (1) The computed incentive is presented to the users. (2a) If the user accepts – The incentive is decreased by a fixed factor and added to the training dataset along with its parameters. (2b) If the user declines – The incentive is increased by a fixed factor and added to the training dataset along with its parameters. (3) The regression tree is re-trained with the extended training dataset.

The incentives are altered before they are added to the dataset to induce a learning process. Accepted incentives can potentially be too high and the algorithm must learn to compute lower incentives. The same goes for declined incentives. In the circumstance when they are too low, the algorithm must learn to compute higher incentives. This will lead to a training dataset that can increasingly adapt to the application case (R6) and also to the machine learning algorithm.

To evaluate the performance of the algorithm and to study the real behavior of users and not just the intended one, we carried out a trial field study for one month (February 2017) in a carsharing system environment. 20 carsharing customers were asked to install a simple user-based relocation smartphone application.

During the field study, 11 users interacted with the various relocation requests. The relocation request within the application contained information regarding the desired relocation, e.g. the distance differences intended and relocation destination as well as the offered incentive, computed by the algorithm itself. During the field study, 20 relocations were requested, of which 11 were accepted and 9 rejected. The average accepted incentive was 2.75 Euro, while various weather conditions were represented.

The temperatures were around 9 degrees Celsius and the average relocation distance was 940m. Given that more than half of the participating users accepted the incentive calculated by the algorithm, it is positively affirmed that the developed algorithm potentially fulfills its original purpose.

4.2 Design Theory

Following Gregor and Hevner [40], each DSR process adds to practice and theory by developing new design theories and by explaining "how to do something". Correspondingly, we developed a design theory for an incentive computation approach for user-based relocation. The developed theory corresponds to the model proposed by Gregor and Jones [38].The design theory is summarized in Table 4.

Table 4. Design Theory of an Incentive Computation Approach for User-Based Relocation (following [38])

Component	*Description*
Purpose and Scope	The approach computes adequate incentives to motivate carsharing users to change their rental destination. Requirements: Self-learning, Context sensitive, Real-time, Cost-efficient, Simple to implement, Adaptive
Constructs	Demand, Supply, Vehicle, User, Relocation, Incentive, Accept, Decline, Destination, Location
Principle of Form and Function	Machine learning (e.g. Regression Tree) can be applied to compute incentives for user-based relocation (R1,R2,R4, R5). The value of declined incentives should be increased and added to the training process (R1, R2, R4). The value of accepted incentives should be decreased and added to the training process (R1, R2, R4). The initial training dataset can be gathered via survey, using weather and bidding as frames/anchors (R1, R2, R5, R6). The trainings parameter should include weather, weekday, temperature, distance, time of the day (R2, R6).
Artifact Mutability	The incentive computation algorithm can be adapted using a context-specific dataset. The input factors as well as the regression tree parameters can be changed.
Testable Propositions	The computed incentives motivate users to comply with user-based relocation. A good incentive remains below employee costs (6€) and motivates certain users (50%)
Justificatory Knowledge	Carsharing Literature, Machine Learning Literature

5 Discussion

The presented study contributes to carsharing literature by developing an incentive computation algorithm to be applied to user-based relocation within carsharing service systems. The computation of incentives to motivate users to relocate a rented vehicle is a fairly new research question [4, 12]. Thereupon, the presented algorithm marks a first step into this new problem domain, proving a valuable base for future research.

Ultimately, the findings of this study have implications for the field of sharing economy research. Carsharing is recognized as a prime example of the sharing economy [2]. However, more research is required to establish the ways in which IS can support carsharing service systems [2]. In this regard, the presented development process and artifacts unravel the contribution of IT and IS to providing sustainable sharing services and creating value with and for customers. To be specific, the focus is currently centralized on how IT and IS help increase the comfort and flexibility of carsharing by implementing user-based relocation. The developed algorithm becomes a key enabler for this concept. It is important to note that the algorithm could also be used for other fields of application e.g. enriching dynamic ride sharing. Passengers could be incentivized to walk to a pickup location by cheaper prices and at the same time drivers need to be incentivized by addition revenue to perform detours. Hence, having a good understanding of how to compute incentives would not only help to improve carsharing but also other mobility services [42–44]. The same reasoning would apply to crowdsource delivery. The algorithm could calculate incentives to motive drivers to perform a detour for a delivery or to walk a certain distance to drop off or pick up a parcel [45].

Additionally, as the presented research demonstrates the different ways in which machine learning methods can be used, it contributes to the research body of machine learning, by introducing it to the domain of carsharing. To the best of our knowledge, no previous study has applied and implemented machine learning models to compute the incentive of user-based relocation. Hence, the successful application and evaluation of the algorithm suggests that machine learning should be further researched in this context.

In context of DSR, this study contributes on several levels. Firstly, it serves as an example for the application of DSR in solving the specific problem of incentive computation in carsharing. Secondly, when positioning the developed artifact within the DSR knowledge contribution framework [40], we argue that the problem of incentive computation for user-based relocation participants is a rather new problem domain, constituted by the current state-of-the-art attention focused on user-based relocation algorithms, without engaging in its real-world applications. Hence, it constitutes a novel problem class. Similarly, the application of machine learning to compute incentives can be seen as a transfer of existing knowledge, classifying the solution maturity as high. Hence, the developed artifact constitutes an exaptation, characterized by a low problem domain maturity and a high solution maturity [40]. Furthermore, the contribution of this study is situated on two different levels of theoretical abstraction [40]. The developed framework and principles of form and function can be characterized as a nascent design theory (level 2), while the implemented and tested algorithm provides a tangible instantiation (level 1).

Besides the theoretical contributions, this study holds valuable implications for practitioners. Firstly, this study demonstrates how to develop and implement an incentive computation algorithm for user-based relocation. Combined with previous studies about the identification of necessary relocations [4, 12], it enables user-based relocation implementation and further practical research within the field. Consequently, carsharing providers are now enabled to substitute operator-based with user-based relocation. With this in mind, the overall cost of relocations can potentially be reduced, increasing the yield of carsharing provider, subsequently making carsharing not only a more sustainable transportation service, but also a more profitable one [12]. Secondly, it underlines how to introduce user-based relocation to a carsharing system. Before implementing any related carsharing projects, carsharing providers must survey their users to train the machine learning algorithm. As the artifact evaluations reveal, it shows potential to compute incentives to motivate users within the system to relocate vehicles.

Regarding any limitations, the two evaluations constitute a limitation of its own. The survey data is only partly representative for the special case of inhabitants of the city in which the carsharing provider is located. Hence, other cities and carsharing systems might find survey data to be insufficient as an initial training dataset. Additionally, the limited scope (20 participants) of the field-test constitutes a limitation. Secondly, in order to get initial results, we applied a research process composed of easy and affordable research and evaluation methods, e.g. expert interviews and small-scale single case application. This process ensures efficiency and iterative adaptation according to any forthcoming interim results. Thus, the next step is to carry out a larger-scaled field-test to ensure generalizability of results. In addition, we research opportunities in identifying and including further parameters in the computation process, such as customer information (e.g. age, rental frequency or attitude to sustainability), to potentially increase the accuracy for individual users. Lastly, during the research process, the participants were questioned directly on the street, following a structured questionnaire to derive a proper incentive. However, willingness-to-pay is difficult to convey in surveys, therefore an alternative approach such as the conjoint analysis has been developed (e.g. [46]). Thus, we see potential for future research in addressing this limitation by applying other methods to validate and expand our knowledge on finding the best incentive/pricing option, providing further input for the incentive computation process.

6 Conclusion

This article presents a self-learning algorithm to compute the necessary incentives for user-based relocation. It bases on a machine learning model and was developed following the DSR framework of Hevner et al. [35]. The algorithm was evaluated within a field study, in which it was revealed that the algorithm had the possibility be used in practice and compute sufficient incentives.

The algorithm enables user-based relocation to be implemented within real-world applications. This will help carsharing providers to improve the flexibility and cost-efficiency of their systems. Furthermore, IS researchers can now study user-based relocation within its intended field of application. In addition to that, with a simple and cost-efficient method, we provide the basis for further extensive and complex methods, so that our solution can be further optimized [35]. Accordingly, the results of this study enable benchmarking, which is an important area of DSR [47].

Overall, our study is in line with the demand to focus on applied Green IS research and the guidelines of vom Brocke [48] to develop tangible research results in the form of IS artifacts that contribute to minimizing the gap between sustainability research and practice, a field currently underrepresented in IS research [49]. Against this background, the facilitation of user-based vehicle relocation in carsharing, through the presented artifact, can be an important building block to further increase the sustainable impact of carsharing.

References

1. Goodall, W., Dovey Fishman, T., Bornstein, J., Bonthron, B.: The rise of mobility as a service: Reshaping how urbanites get around. Delotitte Rev. 111–130 (2017)
2. Remane, G., Nickerson, R.C., Hanelt, A., Tesch, J.F., Kolbe, L.M.: A taxonomy of carsharing business models. In: 37th International Conference on Information Systems. pp. 1–17., Dublin (2016)
3. Clemente, M., Fanti, M.P., Mangini, A.M., Ukovich, W.: The vehicle relocation problem in car sharing systems: Modeling and simulation in a Petri net framework. In: Application and Theory of Petri Nets and Concurrency. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J.-M. Colom and J. Desel (eds), pp. 250–269. Springer Berlin Heidelberg, Berlin (2013)
4. Wagner, S., Willing, C., Brandt, T., Neumann, D.: Data Analytics for Location-Based Services: Enabling User-Based Relocation of Carsharing Vehicles. In: 36th International Conference on Information Systems. pp. 279–287., Fort Worth (2015)
5. Weikl, S., Bogenberger, K.: Relocation strategies and algorithms for free-floating car sharing systems. In: 15th International IEEE Conference on Intelligent Transportation Systems. pp. 100–111., Anchorage (2013)
6. Brendel, A.B., Rockenkamm, C., Kolbe, L.M.: Generating Rental Data for Car Sharing Relocation Simulations on the Example of Station-Based One-Way Car Sharing. In: Proceedings of the 50th Hawaii International Conference on System Sciences. pp. 1–10 (2017)
7. Balac, M., Ciari, F.: Modeling station-based carsharing in Switzerland. In: 14th Swiss Transport Research Conference. pp. 1–16., Monte Verità / Ascona (2014)
8. Ciari, F., Dobler, C., Axhausen, K.W.: Modelling one-way shared vehicle systems - An agent-based approach. In: 13th International Conference on Travel Behaviour Research. pp. 1–14., Toronto (2012)
9. Jorge, D., Correia, G.: Carsharing systems demand estimation and defined operations: A literature review. Eur. J. Transp. Infrastruct. Res. 13, 201–220 (2013)
10. Lee, J., Park, G.L.: Cluster-based vehicle redistribution scheme based on genetic algorithms for electric vehicle sharing systems. Int. J. Softw. Eng. its Appl. 8, 147–158 (2014)
11. Shaheen, S.A., Mallery, M.A., Kingsley, K.J.: Personal vehicle sharing services in North America. Res. Transp. Bus. Manag. 3, 71–81 (2012)
12. Brendel, A.B., Brauer, B., Hildebrandt, B.: Toward User-Based Relocation Information Systems in Station - Based One-Way Car Sharing. In: Twenty-second Americas Conference on Information Systems. pp. 1–10. ,San Diego (2016)
13. Schulte, F., Voß, S.: Decision Support for Environmental-friendly Vehicle Relocations in Free- Floating Car Sharing Systems: The Case of Car2go. In: 7th Industrial Product-Service Systems Conference - PSS, ndustry transformation for sustainability and business. pp. 275–280., (2015)
14. Boyaci, B., Zografos, K.G., Geroliminis, N.: An optimization framework for the development of efficient one-way car-sharing systems. Eur. J. Oper. Res. 240, 718–733 (2014)
15. Barth, M., Todd, M.: Simulation model performance analysis of a multiple station shared vehicle system. Transp. Res. Part C. 7, 237–259 (1999)
16. Brendel, A.B., Brennecke, J.T., Zapadka, P., Kolbe, L.M.: A Decision Support System for Computation of Carsharing Pricing Areas and its Influence on Vehicle Distribution. In: 38th International Conference on Information Systems. pp. 1–21., Seoul (2017)

17. Angelopoulos, A., Gavalas, D., Konstantopoulos, C., Kypriadis, D., Pantziou, G.: Incentivization Schemes for Vehicle Allocation in One-Way Vehicle Sharing Systems. In: IEEE International Smart Cities Conference (2016)
18. Webster, J., Watson, R.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, xiii–xxiii (2002)
19. Banerjee, S., Freund, D., Lykouris, T.: Pricing and Optimization in Shared Vehicle Systems: An Approximation Framework. (2016)
20. Barth, M., Todd, M., Xue, L.: User-Based Vehicle Relocation Techniques for Multiple-Station Shared-Use Vehicle Systems. In: Transportation Research Board 83th Annual Meeting. pp. 1–16., Washington (2004)
21. Bianchessi, A.G., Formentin, S., Savaresi, S.M.: Active fleet balancing in vehicle sharing systems via Feedback Dynamic Pricing. In: Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (Netherlands). pp. 1619–1624., The Hague (2013)
22. Brandstätter, G., Gambella, C., Leitner, M., Masini, F., Malaguti, E., Puchinger, J., Ruthmair, M., Vigo, D.: Dyn. Perspect. on Managerial Decision Making. 22, (2016)
23. Cepolina, E.M., Farina, A.: A new shared vehicle system for urban areas. Transp. Res. Part C Emerg. Technol. 21, 230–243 (2012)
24. Cepolina, E.M., Farina, A.: Urban car sharing:An overview of relocation strategies. WIT Trans. Built Environ. 128, 419–431 (2012)
25. Chow, Y., Yu, J.Y.: Real-time Bidding based Vehicle Sharing. 4–8 (2015).
26. Clemente, M., Fanti, M.P., Iacobellis, G., Nolich, M., Ukovich, W.: A Decision Support System for User-Based Vehicle Relocation in Car Sharing Systems. IEEE Trans. Syst. Man, Cybern. Syst. 1–14 (2017)
27. Di Febbraro, A., Sacco, N., Saeednia, M.: One-Way Carsharing. Transp. Res. Rec. J. Transp. Res. Board. 2319, 113–120 (2012)
28. Gavalas, D., Konstantopoulos, C., Pantziou, G.: Design and Management of Vehicle Sharing Systems: A Survey of Algorithmic Approaches. (2015)
29. Herrmann, S., Schulte, F., Voß, S.: Increasing Acceptance of Free-Floating Car Sharing Systems Using Smart Relocation Strategies : A Survey Based Study of car2go Hamburg. (2015)
30. Jorge, D., Molnar, G., Homen, G., de Almeida Correia, G.H.: Trip pricing of one-way station-based carsharing networks with zone and time of day price variations. Transportation Research Part B: Methodological. 81, 461–482 (2015)
31. Laarabi, H.M., Boldrini, C., Bruno, R., Porter, H., Davidson, P.: On the Performance of a One-way Car Sharing System in Suburban Areas: A Real-world Use Case. Proc. 3rd Int. Conf. Veh. Technol. Intell. Transp. Syst. 102–110 (2017)
32. Wang, H., Cheu, R., Lee, D.H.: Dynamic relocating vehicle resources using a microscopic traffic simulation model for carsharing services. In: 3rd International Joint Conference on Computational Sciences and Optimization, CSO 2010: Theoretical Development and Engineering Practice. pp. 108–111 (2010)
33. Waserhole, A., Jost, V., Brauner, N.: Pricing techniques for self regulation in Vehicle Sharing Systems. Electron. Notes Discret. Math. 41, 149–156 (2013)
34. Weikl, S., Bogenberger, K.: Relocation Strategies and Algorithms for Free-Floating Car Sharing Systems. IEEE Intell. Transp. Syst. Mag. 5, 100–111 (2013)
35. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MISQ. 28, 75–105 (2004)
36. Hevner, A.R.: A Three Cycle View of Design Science Research. Scand. J. Inf. Syst. 19, 87–92 (2007)

37. Gregory, R.W., Muntermann, J.: Heuristic theorizing: Proactively generating design theories. Inf. Syst. Res. 25, 639–653 (2014)
38. Gregor, S., Jones, D.: The Anatomy of a Design Theory. J. Assoc. Inf. Syst. 8, 312–335 (2007)
39. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)
40. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science - Types of Knowledge in Design Science Research. MISQ. 37, 337–355 (2013)
41. Mitchell, T.M.: Machine Learning. McGraw Hill Companies, Boston (1997)
42. Schreieck, M., Safetli, H., Siddiqui, S.A., Pflügler, C., Wiesche, M., Krcmar, H.: A Matching Algorithm for Dynamic Ridesharing. Transp. Res. Procedia. 19, 272–285 (2016)
43. Mirisaee, S.H., Brereton, M., Roe, P.: Bridging the representation and interaction challenges of mobile context- aware computing: designing agile ridesharing. Proc. 23rd Aust. Comput. Interact. Conf. - OzCHI '11. 221–224 (2011)
44. Brereton, M., Ghelawat, S.: Designing for participation in local social ridesharing networks. In: Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10. p. 199., Sydney (2010)
45. Soto Setzke, D., Pfluegler, C., Schreieck, M., Fröhlich, S., Wiesche, M., Krcmar, H.: Matching Drivers and Transportation Requests in Crowdsourced Delivery Systems. In: Twenty-third Americas Conference on Information Systems. pp. 1–10. , Boston (2017)
46. Siegfried, N., Koch, O., Benlian, A.: Drivers of App Installation Likelihood – A Conjoint Analysis of Quality Signals in Mobile Ecosystems. In: 36th International Conference on Information Systems. pp. 1–18., Fort Worth (2015)
47. Rai, A., Burton-Jones, A., Chen, H., Gupta, A., Hevner, A.R., Ketter, W., Parsons, J., Rao, H.R., Sarkar, S., Yoo, Y.: Diversity of Design Science Research. MIS Q. 41, iii–xviii (2017)
48. Vom Brocke, J., Watson, R.T., Dwyer, C., Elliot, S., Melville, N.: Green information systems: Directives for the IS discipline. Commun. Assoc. Inf. Syst. 33, 509–520 (2013)
49. Gholami, R., Watson, R.T.R.T., Hasan, H., Molla, A., Bjorn-andersen, N., Hasan, H., Bjørn-Andersen, N.: Information Systems Solutions for Environmental Sustainability: How Can We Do More? J. Assoc. Inf. Syst. 17, 521–516 (2016)

Sustainability's Coming Home:
Preliminary Design Principles for the Sustainable Smart District

Robert Keller[2,3], Felix Röhrich[1,2,3], Lukas Schmidt[1,3], and Gilbert Fridgen[1,2]

[1]Universität Bayreuth, FIM Research Center, Bayreuth, Germany
[2]Fraunhofer FIT, Project Group Business & Information Systems Engineering, Augsburg, Germany
[3]Universität Augsburg, FIM Research Center, Augsburg, Germany

{robert.keller, felix.roehrich, lukas.schmidt, gilbert.fridgen}@fim-rc.de

Abstract. Consumer trends like local consumption, sharing of property, and environmental awareness change our habits and thereby our surroundings. These trends have their origin in our direct environment, in the districts of our city or community, where we live and socialize. Cities and districts are changing to "smart cities" and "smart districts" as a part of the ongoing digitalization. These changes offer the possibility to entrench the idea of sustainability and build a platform-based ecosystem for a sustainable smart district. This research aims to identify guidelines in form of preliminary design principles for sustainable smart districts. To achieve this, we conduct a structured literature review. On this basis, we derive and develop preliminary design principles with the help of semi-structured interviews and a non-representative sample of the German population. The resulting nine preliminary design principles describe a first insight into the design of sustainable smart districts.

Keywords: Sustainability, Smart District, Platform-based Ecosystem, Smart City, Design Principles

1 Introduction

Sustainability is a "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [1]. The goal of a sustainable development unites the world population more than any other goal in the past [2]. One concept to face sustainability in the area of living are smart cities. These smart city concepts foster sustainable development and face current challenges of our society like immigration, demographic change, and environmental pollution by the means of technology [3–5]. This is one of many reasons why the term "smart city" gained much relevance in the last years [6, 7]. A smart city is able to provide a connected and sophisticated infrastructure to foster economic, ecologic, social, and cultural matters [6] as well as a social-technical view [4]. Current consumer trends with focus on local markets, sharing, mobility, living, and environmental awareness

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

influence the design of the smart cities and its contribution to sustainability [8–11]. Many companies like Uber or nextdoor already recognized the possibilities digital platforms offer in this context.

Since smart cities are very complex systems, it is much easier to plan them with a greenfield approach than on existing cities. To solve this problem, we reduce the complexity to a district perspective by focusing on smaller parts of existing cities. It is also possible to use this district idea for smaller municipalities like small towns, villages, or rural areas in general [12]. Furthermore, consumer trends often take place on a district level and contribute to the ecological, social, and economic development of the district. The implementation of sustainability in districts due to new mobility concepts, sharing concepts, and platform-based collaboration is already happening in many districts [9, 10]. Another example is the use of renewable energy sources for the electricity and heating demand of the district [13, 14]. However, to reach these benefits, there is a need for an adequate information and communications technology (ICT) infrastructure. This technological basis and the connection of all stakeholders lead to smart districts. There is no definition of smart districts in current research. Thus, we define a smart district, based on several smart city definitions and a statement on smart districts by the Smart Cities Information System (SCIS) initiative [7, 15, 16] as follows:

A smart district is a district performing in a forward-looking way in economy, people, governance, mobility, environment, energy, and living, built on a sophisticated, smart ICT infrastructure that ensures benefits for every stakeholder, in particular a high quality of life for every citizen.

The implementation of digital technologies in districts as well as a sustainable mindset, may lead to an ecosystem of a *sustainable* smart district (SSD). Since the term "ecosystem" has become pervasive over the past 20 years, it must be clearly characterized and classified for every research project [17]. For this, we see the SSD as an "ecosystem-as-affiliation" [17]. We consider a platform as the core, the SSD is affiliated with [18, 19]. Due to the central role of the platform, the wording "platform-based ecosystem" (PBE) gains relevance [18].
There are different kinds of platforms [20], which can be the core of a PBE. We see the platform in the SSD context as a digital platform from a sociotechnical point of view [20]. The definition contains not just soft- and hardware, but also organizational processes and standards [21]. Digital platforms stress the idea of modularity [22–24] especially regarding the peripheral components [25]. As Helfat and Raubitschek [18] state, digital platforms are often multi-sided platforms. Multi-sided platforms can be seen as markets that enable direct transactions among several customer groups, with strong network effects between these groups [26, 27]. Multi-sided platforms are of particular interest for us, since in the SSDs PBE many customer groups interact with each other and the value to one party depends on the number and quality of parties on the other sides of the multi-sided platforms [18].

PBEs are an emerging topic that demands extensive research. De Reuver et al. [20] set a research agenda for digital platforms and their surrounding ecosystems. Especially the question on '*how digital platform providers jointly shape platforms with other* *stakeholders*' is of particular interest for the present paper. However, there can be multiple stakeholders trying to influence the design of the PBE, and thus, the digital platform of the SSD and not necessarily one single platform provider. Therefore, we contribute to this scientific discourse by exemplarily analyzing one application: the PBE of the SSD.

Current literature is mostly dealing with sustainable smart city concepts [28–30]. This research often lacks a fine-grained district perspective. Since existing cities are very complex systems it could help transforming them piece by piece, using a district perspective. We identified various projects concerning smart districts like the district project “Smart District Gnigl” as part of the “Smart City Salzburg”. The district project focuses on network and platform effects for local heating, mobility, and education within the district. Furthermore, we also consider smart districts in towns, rural areas, or as new village [12]. Examples for this are the “Smart District Mödlingen” or the “Steimker Gärten”. However, there is little research tackling the adequate implementation and the theoretical background of the SSD. In times of global warming and social alienation, the guidance to a sustainable design is essential for the future smart district. Appropriate guidelines are necessary for understanding how smart districts become SSDs. In this paper we present a first step in this direction and build a theoretical foundation for the implementation of SSDs. We answer the following research question to identify preliminary guidelines for the SSDs PBE.

What are preliminary design principles for a platform-based ecosystem of a sustainable smart district?

2 Methodical Approach

Design Principles (DPs) are guidelines for building design artifacts within design science research [31]. The objectives of such design artifacts are to solve current problems [31] and provide an adequate level of novelty and utility [32]. Generalized prescriptions in the form of DPs aim at extending current design knowledge within this research domain [32]. Thus, DPs are recipes and guidelines for building or describing a specific artifact [33, 34], giving guidance on how to generate a new instance of a class of artifacts [35]. The DPs are usually derived from evaluating actual instances or more abstract conceptualizations [33, 34].

According to Hevner et al. [31], there are four possibilities for building artifacts: constructs, models, methods, and instantiations. For the PBE of the SSD the real-world instantiation is the most reasonable, because instantiations help researchers best to learn about the performance of the particular artifact in the real world [31].

Since there are not many realized smart districts and no SSDs, it is not possible to analyze an existing SSD PBE in detail. Before building an SSD PBE on our own, it is reasonable to first derive preliminary DPs (PDPs). Since there is not much research to SSDs yet, we created a first draft of these PDPs by deducing relevant content from related fields like smart cities with a structured literature review based on Webster and Watson [36] and Fettke [37]. Then, we further developed these PDPs in an iterative process with semi-structured expert interviews based on Myers and Newman [38] and Schultze and Avital [39]. Subsequently, we interviewed possible future inhabitants to gain insights from an additional perspective.

Our research is inspired by the design science cycles of Hevner [40]. In this paper, we were guided by the first rigor cycle by first conducting the structured literature review and then performing semi-structured interviews. The development of the PDPs is a deductive and conceptual process. On the one hand, our PDPs should incorporate already existing knowledge regarding SSDs [31]. Since there is not much research specifically to SSDs, our focus during the structured literature is deriving basic guidelines from related areas like smart city research. On the other hand, we follow a conceptual approach when grouping, formulating and narrowing down the first draft of principles. The interviews also helped us understanding requirements for the SSD, in terms of the relevance cycle. By focusing on the build cycle and deriving PDPs through a iterative process [32, 40], we contribute to the knowledge base on smart districts and form the basis for further operationalization of SSDs [34].

The evaluation framework by Sonnenberg and vom Brocke [41], consisting of four evaluation steps, inspired our research. During this process, we focused on their second step, to iteratively validate and justify our PDPs. Because this step should also encompass the stakeholders of the design artifact we have chosen expert interviews. In each of these interviews we received feedback regarding "ease of use", "efficiency", "generality", "operationality", "completeness", "elegance", meaning the language and structure of the formulations, "simplicity", and "understandability" [31, 32, 35, 41].

In order to contribute to design knowledge, we should ensure an appropriate level of applicability [31]. Therefore, we formulate our nine PDPs in a specific way, as described by March and Smith [32]. To guarantee a high degree of utility and efficiency, we follow the recommendations for precise formulations by Chandra et al. [33]. To consider these recommendations, we formulate our PDPs following the subsequent structure: "Provide the system with [material property – in terms of form and function] to [activity of user / group of users – in terms of action] […]." [33]. Therefore, our PDPs respond to efficiency requirements and maintain a consistent formulation [32].

3 Implementation of our Methodical Approach

To derive our PDPs for an SSD from current literature we conducted a structured literature review based on Webster and Watson [36] and Fettke [37]. We focus on inferring additional knowledge by the means of logical-deductive argumentation. For approaching the literature search in a systematic manner, we codified three search strings for three different research directions (Table 1). Then we evaluated the literature we found. Due to our focus on sustainability, we added this key word to every search string. The same applies to "ecosystem" and synonyms of "district". "smart" is the main prefix describing new digital concepts, no matter whether you look at "smart city", "smart home" or "smart living".

Table 1: Search Strings

Field	*Search string*
Energy and mobility	smart AND (sustainab* OR ecosystem) AND (city OR district OR town OR residential) AND energy AND (mobil* OR flexib* OR local)
Consumer trends	smart AND (sustainab* OR ecosystem) AND (city OR district OR town) AND (consum* OR local OR "sharing economy" OR governance OR residential)
Multi-sided platform	smart AND (sustainab* OR ecosystem) AND (city OR district OR town OR residential) AND ("multi-sided" OR platform)

We thoroughly selected the source material for the literature research. Webster and Watson [36] suggest starting the search within leading journals of the research field. "Sustainable Cities and Society", "Cities" or the journals included in "Senior Scholars' Basket of Journals" could represent important journals for our studies. Since the databases of the main academic publishing houses like SpringerLink or Elsevier contain each of these selected journals, we used these databases in combination with few others to also consider literature from other disciplines [36].

Our search-strings lead us to 4.076 papers, which we evaluated. After screening the titles, 265 publications were left and after reading the abstracts, 95 publications remained. Using those papers, we performed a backward and forward research.

To enhance our first draft of PDPs, which was solely based on the literature we found, we conducted eight expert interviews. We once interviewed three experts at the same time (E4, E5, and E6), and the remaining experts separately (Table 2). We recorded all interviews, which lasted approximately 60 to 90 minutes.

Table 2. Interviewees: Experts

	Business Domain	*Interviewee*	*Employees*
E1	Research Institute	Research Assistant	> 100
E2	Research Institute	Research Assistant	> 100
E3	Real Estate Management	Head of Fund Management	> 500
E4	Engineering Office	Managing Director	> 250
E5	Engineering Office	Consulting Engineer	> 250
E6	Engineering Office	Consulting Engineer	> 250
E7	Research Institute	Research Assistant	> 100
E8	Sustainable City Development	City Planer	> 50
E9	Research Institute	Business Development Manager	> 25.000
E10	Research Institute	Research Assistant	> 100

Inspired by the *design as a search process* of Hevner et al. [31], in the following paragraph, we first illustrate how we developed the PDPs. Subsequently, we present our resulting PDPs in Section 4.

We received relevant feedback in form of practical insights and improvements for the PDPs from the interviews. The feedback is the basis for changes of the PDPs, regarding the selected criteria "ease of use", "efficiency", "generality", "operationality", "completeness", "elegance", "simplicity", and "understandability" [31, 32, 35, 41]. Due to these we adjusted our first draft of PDPs and emphasize important aspects of the SSD. For instance, PDP7 was no part of our first draft. In this case, we got advice from research experts that visionary objectives and goals are crucial for the SSD. Following, the criteria of completeness was not fulfilled. Therefore, we revised relevant literature and formulated a corresponding PDP. One further adjustment was the division of a general IT PDP to PDP3 and PDP5 to reach the criteria "generality". However, some of the comments were contrary to literature or other experts. For example, E7 recommended the exclusionary use of the term "stakeholders" in all PDPs. We followed this recommendation except for PDP5, where we kept the term "user" instead of "stakeholder" in favor of the criteria "understandability". Most feedback faced the criteria "understandability" and "elegance". For example, we reformulated PDP2 after feedback from E4, E5, and E6, to make clear, that stakeholders can be physically outside of the district. With respect to "ease of use" and "operationality", our interview partners gave us valuable recommendations with respect to the implementation in practice. We mitigate PDP9 because E4 noted that complete legal certainty is not reachable in the scope of current legal processes. The demand for changes by the experts shrank from interview to interview. We conducted the major structural and content related changes during the first four interviews. In the last three interviews we did not conduct major changes in the PDPs. In these interviews the experts mainly confirmed the PDPs, sometimes with notes regarding "elegance" and "understandability". Due to their confirmation, we think that our PDPs reached a solid status [42].

We also started with an ex-post validation. Therefore, we conducted interviews with possible future inhabitants. Hence, an appropriate sample should include different groups of society to avoid a possible elite bias [43]. Patton [44] states that under the conditions of academic work such an evaluation is nearly impossible to reach. Therefore, each research should determine a minimum sample size [44]. We talked to 17 potential residents with diverse characteristics (Table 3).

Table 3. Interviewees: Sample of German population

Sex		*Age*		*Education*		*Gross Income*		*Household Size*	
female	8	18-29	6	academic	7	< 15.000 €	7	1 person	6
male	9	30-60	6	non-academic	10	15.000-60.000 €	5	2 persons	6
		> 60	5			> 60.000 €	5	>2 persons	5

4 Results: Preliminary Design Principles for the SSD

In this section we give an overview of our nine PDPs for the PBE of the SSD, which we derived through an iterative process [31, 40] based on a structured literature [36, 37] review and several semi-structured interviews [38, 39].

PDP1: Define adaptable spatial boundaries of the SSD to be able to identify the given characteristics and properties of the SSD.

"Spatial boundaries" stand for a physical connected area of an SSD. The spatial delineation impacts the quality of the SSD out of an environmental, social, and economic perspective [45]. A suitable layout of the SSD fosters sustainable projects and increases the quality of life within it [45]. "Characteristics" describe the geographical location as well as price levels and other intangible attributes. One of the most important distinctions for each SSD project is the difference between districts in cities and districts in rather rural areas. “Properties” stand for different kinds of buildings or public places in the SSD. The SSD should be mixed in its types of buildings, inhabitants, and institutions [45]. This also extends to the importance of mixing social classes and people with different backgrounds. This contributes to sustainability mostly in strengthening the social structures between the inhabitants. Mixing also leads to local labor and consumption within the SSD, which concludes in less transport efforts and more social cohesion. Many experts also emphasized, that these boundaries must be adaptable in case of a changing environment. Especially E9 and E10 stated that there are many different characteristics which can define the boundaries, like buildings or the electric grid.

PDP2: Identify stakeholders taking part in the PBE of the SSD to assign roles.

Hollands [6] concludes, that smart cities start from the human side. In the same way it holds for SSDs. Because of this, every SSD needs to aim for increasing the quality of life of their citizens [46]. For this Cacho et al. [46] suggest the identification of the different stakeholders as one of the first compulsory steps. According to Kennon et al. [47] and E10, the classification is important for designing the digital infrastructure in a suitable way. The identification of all relevant stakeholders is the foundation for cooperation and coordination of investments [48]. Since we focus on the PBE of SSDs, the producers, consumers, providers, and the owner are the basis, which should exist in every SSD [19]. Nevertheless, there are many more different roles in the SSD context [46, 48, 49]. One crucial role is the leader role [18, 50, 51]. This PDP contributes to sustainability by forming the foundation for a social sustainable togetherness. On the other side it also supports economical sustainability by helping to understand the needs of specific stakeholders.

PDP3: Provide the SSD with an adaptable and scalable digital infrastructure to integrate heterogeneous, connected IT systems and features, to facilitate the PBE.

Like smart cities, SSDs depend on the correct and meaningful applications of digital technologies like open data [52], large-scale distributed systems [53], internet of things (IoT) [54], cloud, and fog computing [55] to everyday life [4]. Since everybody in the SSD context should be able to easily design, develop, execute, and share content [49, 56], we consider a central digital platform as core of the whole ecosystem. Hence, we use the term PBE in this context [18]. One possible application of this platform is the field of energy. It can interconnect several energy and legacy providers, virtual power plants, and households [57]. The digital platform can automate network management, enable peer-to-peer energy sharing, minimize operating costs, lower the emission of greenhouse gases, control the electrical energy production as well as the electric vehicles charge/ discharge, and implement demand side integration [13, 58].

Due to the penetration of digital services through all areas of life, the SSD generates a vast amount of data [59]. Big data analysis can improve the performance of SSDs in many areas like energy or mobility, by accomplishing trend analyzes, forecasts or demand planning [59]. According to Cacho et al. [46], it is also possible to generate data from citizens by using social media. As Cacho et al. [46] and E10 stated, the SSD should use the ICT infrastructure and the platform primary as enabler to improve the life of citizens. This can be a chance, especially for SSDs in rural areas, to distinguish themselves from other municipalities, which often do not have the digital infrastructure to attract many people. Furthermore, the SSD should enable the development of innovative green solutions and services, to facilitate a more sustainable smart district.

PDP4: Establish a transparent, cooperative and participatory structure to enable collaboration and competition between stakeholders.

Darking et al. [60] as well as Helfat and Raubitschek [18] emphasize designing governance as a crucial issue. The structure of the SSDs PBE should be transparent and open for all stakeholders to foster cooperation and participation [61]. The claim for transparency comprises the political system, processes, services, market conditions, and the digital platform itself [62]. "Cooperative" means that stakeholders should form partnerships among each other and with actors from outside. Public participation in the form of collaborative decision-making empowers inhabitants of the SSD to express their needs [49, 56, 61]. This implicates market competition and collaboration regarding the distribution and control of resources and power within the SSD [62]. Such cooperative partnerships can lead to desirable synergy effects resulting in benefits for all stakeholders [63]. We do not see competition and collaboration as contrary extremes, but as driving forces for value creation within the SSD [64]. A transparent, cooperative, and participatory structure enables sustainable economic growth and fosters social cohesion. This follows from the involvement of people in the SSD processes and the open and integrating structure [50].

PDP5: Design the services of the SSD in a simple and accessible way to integrate all users.

This PDP prescribes that the services in the SSD have to be user-friendly and people-centric to integrate participants of all age and with every level of education [65]. Designing services in a simple way aims at the usability of these services. Hence, participants will use newly developed services within the SSD more frequently, if the offered solution is intuitive [66]. Including the stakeholders of the SSD in the service creation process strengthens the market competition. Hence, it improves the overall quality and price of the offered services [50].

PDP6: Determine tangible and intangible values to derive an incentive-structure and enable the development of value-adding services, to satisfy the stakeholders' needs.

Tangible values like financial gains are measurable, whereas intangible values are difficult to measure directly. An example for an intangible value is the provision of clean air. The values of the SSD lead to an incentive structure for the stakeholders to encourage their participation and value creation in the SSD [51, 67]. This means, the incentive-structure must be that attractive, that people want to collaborate, for example in either offering services for the SSD, or in using them. There is a strong connection between the services and the incentive-structure, because everybody offering a service wants to benefit. An appropriate incentive structure and value-adding services foster sustainability in various ways. For instance, the waste of food can be reduced by different kinds of food sharing concepts, implemented in the SSD. Additionally, there is potential value for local vendors, due to price reductions for surplus stocks or perishable goods. For the consumer it satisfies needs, by being economically beneficial, because he can purchase the food cheaper than in other circumstances.

PDP7: Continuously monitor the SSD and evaluate feedback to achieve or iteratively adapt visionary objectives and goals.

Literature and experts confirmed, that goals are necessary for the SSD, hence, they are a critical element of the PBE [68]. To ensure that all stakeholders work together, the planner should involve all relevant stakeholders in the finding process [69]. According to Slocombe [69], you should place additional visionary objectives on top of the goals, giving the stakeholders an overall vision for the future development of the SSD. These goals and visionary objectives should also target the implementation of sustainability actions in the SSD [28]. The responsible stakeholder groups should also accompany and reflect the process towards the optimal state [69]. We divide this control mechanism into two different elements. The first is rather quantitative monitoring, the second qualitative feedback. With the help of smart meters and similar devices authorities can automatically check goals. Resulting data sets give indications for future improvements [70]. The other approach of getting feedback from stakeholders of the SSD is more qualitative. This leads to better included stakeholders [56]. After monitoring and considering the feedback, it can make sense to adjust the objectives and goals. For this it is necessary to iteratively adapt these (E10).

PDP8: Integrate public and IT security concepts to provide safety for people, public, and private property.

We will reach sustainable development, if every stakeholder feels safe and protected within the SSD. According to Chifor et al. [71], applications or smart objects, which enable a new way of interaction between humans and their environment, have to consider security aspects. Furthermore, the huge amount of data gathered from different application and services is an issue [70]. It is necessary to ensure that the data is provided for the systems when it is needed. But at the same time, the SSD has to consider the privacy concerns of the people [70]. Furthermore, it is important to still be aware of the "traditional" security aspect for private and public property. For instance, countermeasures must be prepared if the energy supply breaks down.

PDP9: Comply with current law and regulations to aim for legal certainty.

The PBE of the SSD offers lots of chances but its value depends to the regulatory environment [58]. One exemplary regulatory aspect within the SSD is the energy market design. There are several laws that determine which designs are allowed and how taxes and fees are distributed [13]. Today, the regulation in most countries does not allow local peer-to-peer energy sharing [13]. If it were allowed, it would be a viable approach to integrate renewable energy sources in an economical way [13]. However, it is crucial for the SSD to meet current regulatory requirements [13] to target legal certainty. According to E4, it is not possible to achieve full legal certainty because the legal texts leave gaps whose interpretation depends on the opinion of a judge.

5 Discussion

We conducted interviews with ten experts from research, business, and the public sector. All our interviewees are originally from Germany. This represents a limitation of the paper. Experts with other cultural backgrounds could gain alternative insights and improve the PDPs. Furthermore, we conducted interviews, just with a small sample of the German population. The input of 17 people is a brief insight and should be pursued with a bigger number of participants to get representative results. With representative results the planning authorities can ensure that research and real-world projects fulfill the needs of potential inhabitants. Further, we did not conduct an empirical survey on the acceptance of the PDPs in the population.

These limitations in combination with statements of experts as well as parts of the German population point out the critical aspects of this research. Some experts emphasized certain arguments as very important for a successful SSD. According to E9, the trade-off between complexity and economic value is one of the most important aspects of the PBE of an SSD. The SSD must be financially viable and attractive for different stakeholders to successfully implement it in real life. Furthermore, E1, E8, E9, and E10 highlighted the importance of the human and their needs as center of attention within the PBE.

The interviews with a sample of the German population generated insights in the perspective of possible inhabitants. Almost all these interviewees stated that they can envision to live in an SSD. The main incentives they see are a more convenient and sustainable living. Especially the sample of the people above the age of 60 also stressed, that an SSD offers possibilities for people that need help in their everyday life. However, this part of the population was most concerned about losing control and responsibility for their lives. A part of the interviewees in the age between 18 and 60 raised questions about the financing of the SSD. The question who pays for the infrastructure is a relevant decision criterion to them. The question about the financing leads us to an even more relevant question: Who is in charge of the SSD. In the smart city context, current research discusses about private as well as public responsibility [63]. The success of the SSD depends significantly on the stakeholder of the district. Are there sufficient incentives for big and small stakeholders to participate? Which and how much stakeholder must be part of the district?

Some opinions about the concept of the SSD are very similar over the whole group of interviewees. For example, the demand for transparency of the PBE and to the same time the fear of data abuse. However, we found conversely opinions regarding the provided services and priorities of the SSD. Because of the different needs of the groups, there could be an increased risk of pooling inhabitants with similar backgrounds. The responsible authorities should counteract this development in an early stage by balancing the incentive structure and the price level of the SSD. Furthermore, it is questionable which influences multi-sided platforms have on the quality of the provided services. Cennamo et al. [72] illustrate that combining services on platforms can have significant influence on performance within the platform. In general, the SSD still contains open questions and potential undetected risks.

For regular cities, towns and rural areas the emergence of SSDs is a big opportunity. District projects like "Smart District Gnigl" from Salzburg or the German "Open District Hub" project promise to gain important insights in practical benefits of SSD projects. Especially the implications on remaining cities are very important for future district projects. It will help to evaluate a piece by piece approach for transforming cities with the help of SSD concepts.

6 Conclusion, Implication, and Further Research

The idea of a sustainable smart district as a platform-based ecosystem is very complex due to its variety of parties and implications. For this young field of research with rare practical applications, we enforce the awareness of sustainability within the concept. Due to that, the sustainable smart district has a need for preliminary design principles as a way of guidance. With these preliminary design principles, we ensure the consideration of ecologic, economic, and social factors in the district. In this paper we derive nine preliminary design principles from literature as a foundation for the design of the platform-based ecosystem of the sustainable smart district. We developed these principles in an iterative approach based on expert interviews. Furthermore, we interviewed a sample of the German population to ensure their acceptance.

There are many potential topics for further research in the sustainable smart district. First, a real-world instantiation enables the real-world evaluation of the preliminary design principles and carries on the research about sustainable smart districts. After carefully analyzing the results of this instantiation, we could finally evolve the preliminary design principles to design principles in the sense of design science research. Subsequently, we are able to evaluate these design principles according to Sonnenberg and vom Brocke [41], through an extension of the interviews on a complete average of the population. Third, experts from other disciplines and countries would gain further relevant insights to the topic. Fourth, a maturity model for the sustainable smart district would help to understand and classify potential districts. This is especially helpful to capture the current level of sustainability or digitalization within the sustainable smart district. We can also think of evaluating specific technologies or trends in the district context. For instance, future research could find out in which way sustainable smart districts could enable energy sharing within one, but also among several district and thus lead to ecological sustainability. To fully unpack the potential of SSDs, one must also investigate its links to emerging technologies such as electro mobility, smart traffic systems, big data analyses, and much more from a district point of view. Additionally, the three components of sustainability (social, environmental, economic) could be examined in detail, to derive specific recommendations to reach sustainability in the smart district. Furthermore, we raised some open questions in the discussion section, which serve as foundation for further research.

Our results represent a necessary foundation for the following applications in real-world and in science. For the realization of sustainable smart districts our guidelines can help to ensure a sustainable and feasible approach. New research projects can push the idea of the sustainable smart district to concrete concepts and prototypes. In this development a strong cooperation between research and practical application ensures a successful implementation. We believe that this paper contains applicable research, which guides real world projects in creating sustainable smart districts and develop districts, rural areas, and cities towards sustainable future living.

References

1. Brundtland, G.H.: Our Common Future: The Report of the World Commission on Environment and Development. Oslo (1987)
2. Glenn, J.C., Gordon, T.J.: The Millennium Project. Technological Forecasting and Social Change 61, 97–208 (1999)
3. Bibri, S.E.: A foundational framework for smart sustainable city development. Theoretical, disciplinary, and discursive dimensions and their synergies. Sustainable Cities and Society 38, 758–794 (2018)
4. Nam, T., Pardo, T.A.: Conceptualizing smart city with dimensions of technology, people, and institutions. In: 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp. 282–291. New York (2011)
5. Brandt, T., Ketter, W., Kolbe, L.M., Neumann, D., Watson, R.T.: Smart Cities and Digitized Urban Management. Business & Information Systems Engineering 60, 193–195 (2018)
6. Hollands, R.: Will the real smart city please stand up? City 12, 303–320 (2008)
7. Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N.: Smart cities. Ranking of European mediumsized cities. Centre of Regional Science, 1–27 (2007)
8. Neirotti, P., Marco, A. de, Cagliano, A.C., Mangano, G., Scorrano, F.: Current trends in Smart City initiatives. Some stylised facts. Cities 38, 25–36 (2014)
9. Hamari, J., Sjöklint, M., Ukkonen, A.: The sharing economy. Why people participate in collaborative consumption. Journal of the Association for Information Science and Technology 67, 2047–2059 (2016)
10. Zervas, G., Proserpio, D., Byers, J.W.: The Rise of the Sharing Economy. Estimating the Impact of Airbnb on the Hotel Industry. Journal of Marketing Research 54, 687–705 (2017)
11. Angus, A.: Top 10 Global Consumer Trends for 2018. Euromonitor International, 1–43 (2018)
12. Hosseini, S., Frank, L., Fridgen, G., Heger, S.: Do Not Forget About Smart Towns. Business & Information Systems Engineering 60, 243–257 (2018)
13. Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., Weinhardt, C.: Designing microgrid energy markets. Applied Energy 210, 870–880 (2018)
14. Hast, A., Syri, S., Lekavičius, V., Galinis, A.: District heating in cities as a part of low-carbon energy system. Energy 152, 627–639 (2018)
15. SCIS: Smart district, https://smartcities-infosystem.eu/low-carbon-technologies/smart-district (Accessed: 24.05.2018)
16. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart Cities in Europe. Journal of Urban Technology 18, 65–82 (2011)
17. Adner, R.: Ecosystem as Structure. Journal of Management 43, 39–58 (2017)
18. Helfat, C.E., Raubitschek, R.S.: Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. Research Policy 47, 1391–1399 (2018)
19. Alstyne, Marshall, W., Geoffrey, G., Parker, Choudary, S.: Pipelines, platforms, and the new rules of strategy. Harvard Business Review 94, 54–62 (2016)
20. Reuver, M. de, Sørensen, C., Basole, R.C.: The digital platform: a research agenda. Journal of Information Technology 33, 124–135 (2018)

21. Tilson, D., Sorensen, C., Lyytinen, K.: Change and Control Paradoxes in Mobile Infrastructure Innovation: The Android and iOS Mobile Operating Systems Cases. In: 45th Hawaii International Conference on System Science, pp. 1324–1333. Maui (2012)
22. Baldwin, C.Y., Clark, K.B.: The power of modularity. The MIT Press, Cambridge, Massachusetts, London (2000)
23. Garud, R. (ed.): Managing in the modular age. Architectures, networks, and organizations. Blackwell, Malden (2003)
24. Sanchez, R., Mahoney, J.T.: Modularity, flexibility, and knowledge management in product and organization design. Strategic Management Journal 17, 63–76 (1996)
25. Baldwin, C.Y., Woodard, C.J.: The Architecture of Platforms: A Unified View. Platforms, markets and innovation, 19–44 (2008)
26. Eisenmann, T., Parker, G., van Alstyne, M.: Platform envelopment. Strategic Management Journal 32, 1270–1285 (2011)
27. Rochet, J., Tirole, J.: Two-sided markets: a progress report. The RAND Journal of Economics 37, 645–667 (2006)
28. Trindade, E.P., Hinnig, M.P.F., da Costa, E.M., Marques, J.S., Bastos, R.C., Yigitcanlar, T.: Sustainable development of smart cities. A systematic review of the literature. Journal of Open Innovation 3, 1–14 (2017)
29. Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., Airaksinen, M.: What are the differences between sustainable and smart cities? Cities 60, 234–245 (2017)
30. Hirst, P., Hummerstone, E., Webb, S., Karlsson, A., Blin, A., Duff, M., Deakin, M.: JESSICA for Smart and Sustainable Cities - Horizontal Study - Final Report. London (2012)
31. Hevner, A., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly 28, 75–105 (2004)
32. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decision Support Systems 15, 251–266 (1995)
33. Chandra, L., Seidel, S., Gregor, S.: Prescriptive Knowledge in IS Research. Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. In: 48th Hawaii International Conference on System Sciences, pp. 4039–4048. Kauai (2015)
34. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37, 337–355 (2013)
35. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
36. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26, xiii–xxiii (2002)
37. Fettke, P.: State-of-the-Art des State-of-the-Art. Wirtschaftsinformatik 48, 257–266 (2006)
38. Myers, M.D., Newman, M.: The qualitative interview in IS research. Examining the craft. Information and Organization 17, 2–26 (2007)
39. Schultze, U., Avital, M.: Designing interviews to generate rich data for information systems research. Information and Organization 21, 1–16 (2011)
40. Hevner, A.R.: A Three Cycle View of Design Science Research. Scandinavian journal of information systems 19, 87–92 (2007)
41. Sonnenberg, C., Vom Brocke, J.: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Proceedings of the 7th International Conference on Design Science Research in Information Systems, pp. 381–397. Las Vegas (2012)
42. Marshall, B., Cardon, P., Poddar, A., Fontenot, R.: Does Sample Size Matter in Qualitative Research? A Review of Qualitative Interviews in IS Research. Journal of Computer Information Systems 54, 11–22 (2015)

43. Heiskanen, A., Newman, M.: Bridging the gap between information systems research and practice: the reflective practitioner as a researcher. In: Proceedings of the 18th International Conference on Information systems, pp. 121–132. Atlanta (1997)
44. Patton, M.Q.: Qualitative research & evaluation methods. Integrating theory and practice. Sage, Thousand Oaks (2015)
45. Meijer, M., Adriaens, F., van der Linden, O., Schik, W.: A next step for sustainable urban design in the Netherlands. Cities 28, 536–544 (2011)
46. Cacho, N., Lopes, F., Cavalcante, E., Santos, I.: A smart city initiative. The case of Natal. In: Proceedings of the IEEE International Smart Cities Conference, pp. 1–7. Trento (2016)
47. Kennon, N., Howden, P., Hartley, M.: Who really matters?: A stakeholder analysis tool. Extension Farming Systems Journal 5, 9–17 (2009)
48. Ielite, I., Olevsky, G., Safiulins, T.: Identification and prioritization of stakeholders in the planning process of sustainable development of the smart city. In: 7th International Conference on Intelligent Computing and Information Systems, pp. 251–257. Cairo (2015)
49. Aguilera, U., Peña, O., Belmonte, O., López-de-Ipiña, D.: Citizen-centric data services for smarter cities. Future Generation Computer Systems 76, 234–247 (2017)
50. Fatemeh, N., Itälä, T., Reuver, M. de: Collective Action in a Smart Living Platform Ecosystem. The Role of Platform Leadership and Platform Openness. International Conference on Mobile Business, 10–21. Delft (2012)
51. Bianco, W.T., Bates, R.H.: Cooperation by Design. Leadership, Structure, and Collective Dilemmas. The American Political Science Review 84, 133–147 (1990)
52. Ojo, A., Curry, E., Zeleti, F.A.: A Tale of Open Data Innovations in Five Smart Cities. In: 48th Hawaii International Conference on System Sciences, pp. 2326–2335. Kauai (2015)
53. Steen, M., Pierre, G., Voulgaris, S.: Challenges in very large distributed systems. Journal of Internet Services and Applications 3, 59–66 (2012)
54. Silva, R., Sa Silva, J., Boavida, F.: A symbiotic resources sharing IoT platform in the smart cities context. In: 10th International Conference of Intelligent Sensors, Sensor Networks and Information Processing, pp. 1–6. Singapore (2015)
55. Cheng, B., Solmaz, G., Cirillo, F., Kovacs, E., Terasawa, K., Kitazawa, A.: FogFlow. Easy Programming of IoT Services Over Cloud and Edges for Smart Cities. IEEE Internet of Things Journal 5, 696–707 (2018)
56. Lea, R., Blackstock, M., Giang, N., Vogt, D.: Smart cities. Users and Developers to Foster Innovation Ecosystems. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 1535–1542. New York (2015)
57. Karnouskos, S., Holanda, T.N.d.: Simulation of a Smart Grid City with Software Agents. In: Proceedings of the third UKSim European Symposium on Computer Modeling and Simulation, pp. 424–429. Athens (2009)
58. Garau, M., Ghiani, E., Celli, G., Pilo, F.: Tecno-economic and environmental assessment of a full electric smart city eco-district. In: AEIT International Annual Conference, pp. 1–6. Cagliari (2017)
59. Curry, E., Dustdar, S., Sheng, Q.Z., Sheth, A.: Smart cities – enabling services and applications. Journal of Internet Services and Applications 7, 1–3 (2016)
60. Darking, M., Whitley, E.A., Dini, P.: Governing diversity in the digital ecosystem. Communications of the ACM 51, 137–140 (2008)
61. Paulin, A.: Informating Smart Cities Governance? Let Us First Understand the Atoms! Journal of the Knowledge Economy 7, 329–343 (2016)

62. Paulin, A.: Technological Ecosystems' Role in Preventing Neo-Feudalism in Smart-City Informatization. In: Proceedings of the 25th International Conference Companion on World Wide Web, pp. 333–337. New York (2016)
63. Lee, J.H., Hancock, M.G., Hu, M.-C.: Towards an effective framework for building smart cities. Lessons from Seoul and San Francisco. Technological Forecasting and Social Change 89, 80–99 (2014)
64. Lavie, D.: Alliance portfolios and firm performance. A study of value creation and appropriation in the U.S. software industry. Strategic Management Journal 28, 1187–1212 (2007)
65. Ding, X., Zhong, W., Shearmur, R.G., Zhang, X., Huisingh, D.: An inclusive model for assessing the sustainability of cities in developing countries – Trinity of Cities' Sustainability from Spatial, Logical and Time Dimensions. Journal of Cleaner Production 109, 62–75 (2015)
66. Vatsikas, S., Kalogridis, G., Lewis, T., Sooriyabandara, M.: The Experience of Using the IES Cities Citizen-Centric IoT Platform. IEEE Communications Magazine 55, 40–47 (2017)
67. Wareham, J., Fox, P.B., Cano Giner, J.L.: Technology Ecosystem Governance. Organization Science 25, 1195–1215 (2014)
68. Gretzel, U., Werthner, H., Koo, C., Lamsfus, C.: Conceptual foundations for understanding smart tourism ecosystems. Computers in Human Behavior 50, 558–563 (2015)
69. Slocombe, D.S.: Defining Goals and Criteria for Ecosystem-Based Management. Environmental Management 22, 483–493 (1998)
70. Al Nuaimi, E., Al Neyadi, H., Mohamed, N., Al-Jaroodi, J.: Applications of big data to smart cities. Journal of Internet Services and Applications 6, 1–15 (2015)
71. Chifor, B.-C., Bica, I., Patriciu, V.-V.: Sensing service architecture for smart cities using social network platforms. Soft Computing 21, 4513–4522 (2017)
72. Cennamo, C., Ozalp, H., Kretschmer, T.: Platform Architecture and Quality Trade-offs of Multihoming Complements. Information Systems Research 29, 461–478 (2018)

Substitution of hazardous chemical substances using Deep Learning and t-SNE

Patrick Lübbecke, Nijat Mehdiyev, Peter Fettke

German Research Center for Artificial Intelligence, Institute for Information Systems, Saarbrücken, Germany
{patrick.luebbecke, nijat.mehdiyev, peter.fettke}@dfki.de

Abstract. Manufacturing companies in the European Union are obliged to regularly analyze their recipes to find safer alternatives for hazardous substances. Unfortunately, available substance information is dispersed, heterogeneous and stored in databases of many private and public entities. In addition, the number of existing chemical substances already surpassed 85,000 with over 200 attributes describing substance characteristics, which makes it impossible for experts to collect and manually review this data. We tackle these issues by introducing a novel machine learning approach for alternative assessment. After developing a central database, we design an approach that performs nearest neighbor search in latent space obtained by deep autoencoders. Furthermore, we implement a post-hoc explanation technique, t-SNE, to visualize deep embeddings that enables to justify model outcomes. The application in a real-world project with a manufacturer shows that this approach can help process experts to identify possible replacement candidates more quickly and fosters comprehensibility through visualization.

Keywords: worker safety, sustainability, alternatives assessment, deep learning, machine learning

1 Introduction

In the member countries of the European Union, many laws have been installed that aim at reducing the impact that industrial processes may have on the environment and the health of workers and consumers. A prominent example is the RoHS directive (*Restriction of Hazardous Substances*) which, among others, prohibited manufacturer from the use of leaded solder and the sale of any products that contain these regulated substances [1]. Apart from the prohibition to use certain substances, which are determined by the government, legislators within the EU impose companies to take individual action towards reducing the environmental or health-related impact of industrial processes.

As part of these regulations, companies must regularly evaluate the possibilities to substitute hazardous substances used in production with less harmful alternatives. Alternatives assessment (AA) is a process for identifying, comparing, and selecting

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

safer alternatives to chemicals of concern (including those in materials, processes, or technologies) on the basis of their hazards, performance, and economic viability [2].

The issue of AA is currently engaged by researchers, government agencies, as well as by NGOs [3, 4]. Government agencies like the European Chemical Agency (ECHA) primarily focus on the management aspect of chemicals by providing data for labeling and classification of substances. Researchers and NGOs, on the other hand, are working on process aspects when conducting AA, resulting in frameworks that support manufacturers in their AA activities [3]. Currently, there are several frameworks for alternatives assessment [3]. At the beginning of an AA process, most researchers suggest the identification of chemicals of concern that will be subject to AA [5–8]. The substances' physio-chemical properties, the human health hazards, and ecotoxicity are subject to evaluation in most of the ten available frameworks identified by [3]. In all frameworks, an economic assessment is part of the AA process after the ecotoxic and health hazard assessment is done [3]. Identifying alternative substances is considered critical in all frameworks, but mostly remains a manual task and is described only on a very vague level. The authors make suggestion on what factors to consider [8], to use market view or literature research for identifying alternative substances [5], or pose questions that should be addressed when trying to find alternative substances [9].

The limitation of available research is threefold. First, most of the existing frameworks are only guidelines for decision making or suited to a certain environment, rather than being a prescriptive protocol to follow. Currently, there is a lack of techniques and information systems (IS) that support companies in finding meaningful alternative substances by addressing some of the issues that come with AA. The second limitation is the complexity of data or the lack of data [3]. The ECHA provides a public database with over 85,000 registered substances on their website [10]. Each data set comes with a classification of the physio-chemical characteristics for each substance using more than 200 different attributes. In addition to the ECHA database, there exist more databases from individual manufacturers (e.g. BASF), trade associations (e.g. GISBAU), governmental agencies, or national social insurance carriers. Since these entities focus on other aspects than ECHA, their data is characterized by other attributes than the ECHA data. This makes it almost impossible for process experts to use this kind of data for AA due to the sheer number of substances and variety of the data, which brings us to the third limitation of current research. Apart from information systems that provide data for individual substances through a web site, there is currently no approach available that helps process experts in extracting and/or condensing meaningful information for their AA activities from existing data.

To the best of our knowledge, there is currently no approach available that specifically supports the identification of finding alternative substances using sophisticated methods from the area of machine learning. We aim at filling this void with the research provided in this paper. We tackle the latter two limitations by designing a software artifact that contains (i) a substance database, which captures the characteristics of the substances that we develop by extracting the data from different sources, (ii) a machine learning (ML) model that examines the substance data to uncover similarities among substances and identifies substitution candidates based on their physio-chemical and hazard-related attributes as well as their functional role [11],

(iii) an explanation interface for ratification of the generated list of substitution candidates. The ML component is based on a nearest neighbor search algorithm combined with the deep stacked autoencoders. For explainability purposes we use the recognized visualization technique, t-SNE, which is assumed to project the obtained autoencoder representations in a two-dimensional space. The resulting IS can potentially help process engineers from different domains to automatically check their bills of materials for the availability of alternative substances that fulfill the same purpose of the original substance, but which are less harmful for the environment, for the workforce, and end users. We demonstrate the applicability of the proposed IS artifact for the production process of a leading German manufacturer of ceramics.

The remainder of the paper is organized as follows: we provide the applied research method in-depth in Section 2. Section 3 introduces the architecture overview and the individual components of the proposed decision support system. More specifically, we explain the database that was created, the ML models that we applied on that data and the explanatory component of the ML technique. Insights on the applicability of our approach are provided in section 4 with a use-case and evaluation in a real company, before we conclude the paper with a summary and a brief overview of future work in section 5.

2 Research Method and Design

The research method applied in this paper shall be characterized as design science research (DSR) [12, 13]. DSR is an important paradigm in IS research and professional practice and serves as a guideline for the process of constructing socio-technical artifacts in the IS domain [14]. The goal of DSR is to solve existing problems with new or improved IS-based solutions [15]. We follow the process of Peffers et al. [13] for creating design science artifacts. The research environment is an interdisciplinary consortium research project that allows us to evaluate our artifacts in a real-world manufacturing situation and collect continuous feedback from project partners [16].

Table 1. Process of creating Design Science Artifacts (adapted from [13])

Step 1: Problem Identification & Motivation	**Step 2: Objectives of a Solution**	**Step 3: Design & Development**	**Step 4: Demonstration**	**Step 5: Evaluation Step 6: Communication**
– Lack of annotated data – Infeasibility of the manual search process – The opaqueness of the black box machine learning techniques	– Collecting data necessary for alternatives assessment – Searching alternative substances with ML techniques – Evaluating the alternatives in terms of ecological and health-related criteria	– Data model and database – ML techniques for reducing complexity of data set – Explanation Interface for ML techniques – Information system for alternative assessment	– Application of the artifacts in the problem context for 4 substances of concern	– Evaluation of the artifacts in a real-world scenario – Presentation of results to academics and to practitioners

At the beginning, the *problem* must be identified and the need for a new or better solution to the problem should be motivated. As mentioned before, the problem in our case is the current fragmentation of substance data in different databases of individual manufacturers and the lack of methods to process this data [3]. In most current applications such as *P2OASys*[1] or *COSHH*[2], the domain experts must seek for alternative substances manually by analyzing the multidimensional application purposes and the hazard profile of each individual substance, which results in high cognitive load and suboptimal results. Automating the search process with ML techniques is a good alternative, which in turn also requires a careful design and the suitable choice of the approaches. Finally, to embed the ML systems to the decision-making processes it is important to establish the trust to their outcomes, which requires making them explainable by considering the context of application and the user requirements.

In step 2 (*objectives*), the goal of the research must be emphasized. The objective of a solution serves as a baseline of knowledge on which to evaluate the novelty of the created artifacts [14]. The objectives of our approach are to tackle the lack of a sophisticated database with annotated substance data. We do this by extracting necessary data from various sources and creating a single database with harmonized data. We then apply ML techniques to provide means for the expert to evaluate and select alternative substances.

Step 3 (*design & development*) is the phase where the artifacts for the intended solution are created. This can be new constructs, models, methods, or instantiations [12]. We create the design of the database and ML models and implement them as a prototype.

The application of the new artifacts in a suitable context is subject to step 4 (*demonstration*). We chose the context of manufacturing in the ceramics industry to apply our prototype for AA and demonstrate the benefits in this domain.

Step 5 (*evaluation*) is where the final evaluation of the artifacts happens. Based on metrics or observations, the results of the artifacts are evaluated regarding their performance in the field. Since expert knowledge in this step is necessary, we included the ceramics manufacturer to evaluate the resulting list of substitution candidates for five substances that the manufacturer is currently using.

The process of DSR concludes with step 6, where the *communication of the results* to research communities and practitioners takes place.

3 Description of the Proposed Decision Support System

3.1 Architecture overview

With our adapted process (see fig. 1), we follow the framework of the National Research Council [17] for AA. The NRC recommends adopting a multi-perspective approach for AA that begins with the identification of substitution candidates. This is probably the most crucial part of AA due to the information overload for experts. We

[1] see https://p2oasys.turi.org/GetStarted/p2oasys.php

[2] see http://www.hse.gov.uk/coshh/essentials/coshh-tool.htm

support this process with our ML technique, which identifies similar substances, according to their attributes. The physio-chemical attributes must then be assessed by a domain expert to find out, which of the candidates are appropriate for the desired application purpose or product. A comparative exposure assessment for human health hazards as well as ecotoxicity should then be conducted with the remaining substitution candidates. This step is supposed to identify the substance with the smallest negative impact on environment or workforce. We support this step with a web application by providing visualization of the related health or environmental risk for the substitute candidates. Depending on the application environment, the process expert can weigh the attributes differently. For example, in a scenario where water is drained off to the environment after usage, a substance with less risk for causing aquatic toxicity would be more appropriate whereas a substance whose dust can be harmful to the workforce would be less problematic in a fully automated process. Finally, economic considerations like market prices or availability of substances are considered. These three steps are recommended by the majority of established AA frameworks [17].

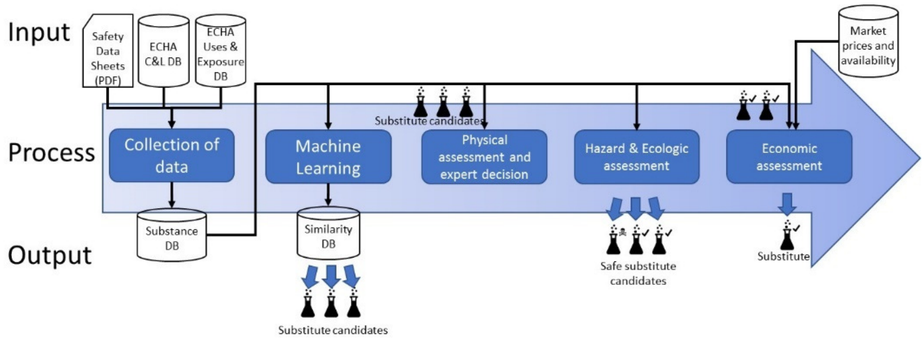


Figure 1. Architecture of the information system

Since the related research project is ongoing and we focus on environmental and health issues in this paper, we present the results up to the hazard & ecologic assessment and focus on the ML component in the following sections.

As mentioned in the introduction, there is currently a need for IS that support manufacturers in comparing available substances, e.g. for planning and maintaining product design. This IS must be capable of analyzing large sets of chemical substance data using ML models. Since many manufacturers have no data science experts available, the results of the ML techniques must be provided in a way that can easily be understood by domain experts.

The solution that we propose is based on a web application that we implemented using Java Enterprise for the user interface and business logic, and R statistical computing programming language for the ML components. In the end user perspective, the user can start a new AA process by providing the CAS# of the substance that is subject to the assessment along with meta data for the AA project. Thereafter, the user can choose how many possible substitutes should be presented and if potentially dangerous substances from the list of substances of very high concern (SVHC) which is maintained by ECHA should be included in the suggestions. The user will then be provided with the interactive scatter chart from Figure 3 with the initial substance highlighted. The samples near the examined instance are provided as colored dots depending on the number of suggestions that was selected. Once the users click these points, these substances will immediately be added to a table below the chart that shows the physio-chemical attributes of every selected substance. After a manual review of the relevant attributes, the user can select the remaining candidates for the next process step. In the next step, the remaining suggestions will be compared among each other regarding their hazard for human and environment. This is done using the GHS column model [18], where the hazard of each substance is benchmarked regarding six different dimensions (see fig. 4).

The administrator of the web app can access the machine learning perspective to start a new learning phase. During the learning, the substance data is retrieved from the database, the autoencoders and t-SNE models are applied to identify and to visualize the nearest neighbors for each substance. The results in the form of an x- and y-value for the position of each substance in a 2d space are stored in the database. The learning must be recomputed as soon as new substances considered are added to the database.

3.2 Database Component: Developing a Central Database

The design of a comprehensive database that can be used for AA in manufacturing companies requires collecting and processing available data. The objective of the ECHA is to collect data about hazardous substances that are imported to or produced in the European Union. ECHA is constantly assessing the chemical-related risks of many substances through manufacturer provided dossiers, own tests, and chemical analyses. Thus, ECHA has available a large volume of data, that is publicly available through their web site.

Most important for identifying initial substitute candidates is data regarding the physical attributes of substances. This kind of data is provided by ECHA through their *registered substances list*, where information about the use and exposure (U&E) of substances is available. The U&E data is classified by the following use descriptors.

Table 2. Attributes of the uses and exposure data set

Attribute type	**Sector of use** (SU, 24 attributes)	**Process category** (PROC, 31 attributes)	**Product category** (PC, 45 attributes)	**Article Category** (AC, 81 attributes)	**Environmental Release Category** (ERC, 26 attributes)
Description	The sector in which the substance is used (e.g. industrial, consumer)	The application techniques or process types	Types of chemical products in which a substance is used.	Type of article in which the substance has been processed.	Broad conditions of use from an environmental perspective.

A second type of data available from ECHA is the *classification and labelling inventory* (C&L). The C&L provides electronic public access to possible hazards for workforce and environment that can be caused by a substance. This data is important for the hazard and ecologic assessment step. The basis for the characterization is the

Globally Harmonized System of Classification and Labelling of Chemicals (GHS) with its 100 attributes in total. These binary-coded attributes describe, whether a substance is causing a certain physical, health, or environmental hazard or not.

As a third source for acquiring substance data we implemented a software tool for extracting relevant information from existing safety data sheets (SDS) in the PDF format. These sheets provide a standardized set of data regarding substance meta data (Manufacturer, retail name), composition/information on ingredients, or relevant identified uses of the substance or mixture. The matching between the data extracted from SDS and the C&L inventory was made through the chemical abstract number (CAS#), a unique identifier for every substance which is available at every SDS and in the C&L and U&E inventory.

Through these three data sources, we could collect 13,133 data instances regarding U&E of substances and 85,845 data instances of the C&L inventory.

3.3 Modelling Component: Nearest Neighbor Search and Deep Autoencoders

3.3.1 Finding Alternative Substances with Nearest Neighbor Search

Once the central database is developed, we aim to propose an intelligence component that presents the production experts the most suitable alternatives to the substances that are used in the current production processes. For this purpose, the most similar matches to high dimensional query vector that captures different aspects of the usage- and hazard profile can be extracted by using the nearest neighbor search algorithm. However, searching the similar items in large datasets is a challenging issue. Hence, the suitability of the retrieved neighbor instances via nearest neighbor search algorithm is affected negatively by the inappropriate adoption of distance/similarity measure and existence of various class irrelevant features [19]. Furthermore, the process of measuring the similarities between the query case and historical items with high feature dimensions from very large validation sets is very time consuming and suffers from the curse of dimensionality which makes infeasible to identify the exact matches in reasonable computational costs [20].

To alleviate these obstacles, various techniques have been proposed and extensively investigated [20]. Neighborhood identification in the latent space obtained by applying dimensionality reduction techniques is considered as the most promising approaches among others [21]. Earlier studies in this field investigated diverse linear approaches such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) among others to map the original high dimensional feature space to the low dimensional latent space linearly [19]. However, such a linear transformation shows an inability to model the higher-order correlation in the original data space and is deficient in capturing the intrinsic class-specific data manifold [19, 22]. On these grounds, we can argue that searching for similar instances in the latent space obtained by linear transformation may result in the retrieval of the inappropriate explanation items.

Different non-linear transformation and dimensionality reduction approaches have been proposed to surmount the problems arising from these shortcomings. These techniques are assumed to place the items having the same characteristics near to each other in the non-linearly transformed low-dimension feature space [19]. In our study, we employ the stacked autoencoder based deep neural networks, which extract features that constructs useful higher-level representations. The empirical experiment results on various datasets by [22, 23] provide confirmatory evidence that using the latent representations obtained by autoencoders shows superior performance compared to linear PCA, LDA and other non-linear approaches such as linear embeddings in retrieving similar items.

3.3.2 Autoencoder Neural Codes for Nearest Neighbor Search

Autoencoders is the type of unsupervised feed-forward neural networks with three layers namely, an input layer, a hidden layer and an output layer in which the training purpose is defined as reproducing the input data at the output layer. The component of autoencoder networks which computes the hidden layer activations from the input data are referred as encoders. The encoder employs a non-linear mapping function to obtain the hidden layer representation. Following this, the decoder component maps the latent representation of the input data obtained and presented in the hidden layer through the encoding process to a reconstructed vector in high-dimensional input space by using the chosen non-linear mapping function with reverse mapping parameter set. The parameter sets of both encoding and decoding layer are optimized simultaneously with the aim to minimize the average reconstruction error.

Deep neural networks can be built by stacking multiple layers of autoencoders which have already been trained locally as described in the previous section [24]. The training process of the stacked autoencoders with k layers is initialized by training first autoencoder in which the original input data are used both in input and output layers. The learned hidden layer activations are then wired to input and reconstruction layers of the second autoencoder with the purpose to obtain the corresponding hidden layer features and parameters. This greedy layer-wise learning is performed until all features including the ones from last k^{th} hidden layer are extracted and all relevant parameters, weights, and biases are initialized. An example training process of deep neural networks with two hidden layers is depicted in Figure 2.

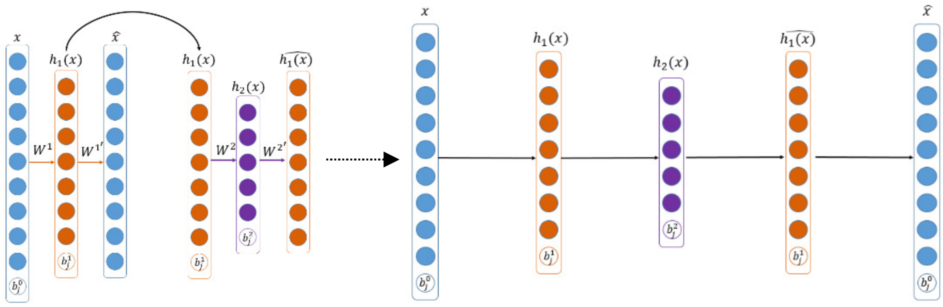


Figure 2. Layer-wise training of stacked autoencoders

As described above the main idea of (stacked) autoencoders is to learn the useful latent representations by teaching the network to copy the input to output. For this purpose, it is very important to add either architectural constraints or employ the regularization techniques during the learning process. In the last decade various autoencoder architectures such as stacked denoising autoencoders, variational autoencoders, contractive autoencoders, regularized autoencoders, undercomplete autoencoders and others have been proposed which are assumed to increase the quality of the extracted latent features. Since we aim in our study also to conduct the nearest neighbor search process based on hidden layer activations faster and more efficiently, it is reasonable to adopt an autoencoder architecture with the decreasing width (layer sizes) of hidden layers, namely undercomplete autoencoders. The main idea behind this type of the autoencoders is to constrain the number of nodes in the hidden layer and to force the model to learn the most important features form the inputs while minimizing the reconstruction error.

After mapping all substances in the dataset onto useful latent space with autoencoders we can find the similar instances to the given query instance by searching its neighbors in the latent space. For this purpose, we first extract the bottleneck features of the query instance by feeding its original input values to the learned deep autoencoder, by computing the hidden layer activations and by calculating the distances between them and the bottleneck features of all other instances. The list of substances with the smallest distance to the query instance is presented to the experts, which are later examined whether they are sustainable alternatives in terms of productions and economic factors.

3.4 Explanation Component: Visualization with t-SNE

In order to exploit the full potential of ML techniques, their outcomes have to be embedded to the production and business processes that creates value by extending the corporate ability to gain new insights [25]. However, a series of recent studies have indicated that the lack of trust in the machine learning model with opaque reasoning mechanisms is considered as one of the main obstacles in operationalizing the data driven analytics that in turn contributes to the extending gap in the scientific developments and their practical applications [26, 27]. Making the advanced black-box models such as deep learning techniques implemented in the current study or their outcomes explainable is considered a potential solutions that has recently received substantial interest [28]. The practicability and reliability of the explanations depend significantly on the context of explanation situation and the target audience that must be considered when choosing and designing the relevant techniques. In our use case, the machine learning explainability is required for the end users, namely the production domain experts, who have little interest in understanding the reasoning procedure of the implemented nearest neighbor search in latent space obtained by deep stacked autoencoders [29]. With this in mind, the post-hoc explanation family techniques are suitable for this purpose since they are assumed to increase the end users' confidence to the ML models by allowing to justify the outcomes delivered by the system [30].

Since the proposed ML model in this study provides the end users the list of similar alternatives to the query case, it already incorporates an intrinsic example-based explanation ability. The domain experts can ratify the suitability of the generated alternatives list by using their domain knowledge and evaluate their validity without a need for examining how and why it was generated by the system. However, such explanations facilitate the users to understand the model locally by providing the list of alternatives for one substance at a time. To enable the users to analyze and understand the entire model outcomes in the whole instance space we adopt another post-hoc explanation technique, visualization with t-Stochastic Neighborhood Embeddings (t-SNE). t-SNE is considered as one the most powerful dimensionality reduction and visualization techniques, that can effectively visualize the high dimensional data by assigning each data point a location in a two or three dimensional map [31]. The main superiority of the t-SNE is its ability to visualize the similarity data by retaining the local structure of the data while also providing the relevant information about the global structure. It is important to note that we visualize the embeddings obtained by the compression of the original data via deep stacked autoencoders with an extension of the original t-SNE technique, namely its implementation with Barnes-Hut approximation algorithm. Implementing the later approach leads to substantial computational efficiency by enabling to embed the high dimensional data with the runtime $O(N \log N)$ instead of $O(N^2)$.

4 Experimental Settings and Results

4.1 Use Case: Substitution of Hazardous Substances in Ceramics Industry

The research covered by this paper was conducted in the ceramic industry. The company that served as partner is a German manufacturer of sanitary ceramics with manufacturing facilities in different European countries. The company uses a variety of chemical substances and mixtures during the three steps of manufacturing. At the very beginning, substances are used, for example, as adhesive or cleansing agent for the construction of the molding tools (i.e. the negative). Then the compound for molding is mixed in a large pug mill from several substances. After the mix is ready, the molding tool is used for high-pressure injection molding of the mix which forms the workpiece. In a series of further manual steps, grinding and glazing of the workpiece takes place. In each of these steps, workers can be exposed to hazardous substances through skin contact or inhale of dust. Most substances have a wide range of short and long-term effects, ranging from skin irritation to carcinogenicity. To reduce the risk for workforce and environment, the firm is looking for new techniques of AA that are robust and can cope with the challenges that massive amount of data present today.

4.2 Tools and Model Parameters

All models for machine learning and explanation components of the proposed solution were developed by using the R statistical computing programming language. The stacked autoencoders were implemented on top of the h2o package [32]. The input layer consists of 208 nodes and the bottleneck code layer size was defined as 20. To avoid the overfitting, we adopted the early stopping technique in which the Mean Squared Error was used as stopping criterion. All layers of the implemented stacked autoencoders used the Tanh activation function. For nearest neighbor search we used the brute force nearest neighbor search in the latent space obtained by stacked autoencoders by computing the Euclidean distance between the query case and all other substances in the dataset. For this purpose we used the FNN package [33]. Performing the nearest neighbor search in the latent space has resulted in an acceleration by factor 11 compared to the search in original high dimensional space.

Table 3. t-SNE and stacked autoencoder parameters

t-SNE		Autoencoder	
Parameter	Value	Parameter	Value
Dimension of Embedded Space	2	Layers	208-50-20-50-208
Learning Rate	200	Activation	Tanh
Perplexity	30	Epochs	1000
Iterations	1000	Learning Rate	0.005
Initial Momentum	0.5	Rate Annealing	0.000001
Momentum Switch Iteration	250	Initial Momentum	0.5
Final Momentum	0.8	Final Momentum	0.99
Theta (Speed/Accuracy Trade-off Parameter)	0.5	Rate Decay	1
Early Exaggeration	12	Initial Weight Distribution	UniformAdaptive

Following this, we visualized all data represented in the bottleneck neural codes of the implemented autoencoder by using the t-SNE with Barnes-Hut approximation algorithm and follow the training instructions by its developers to obtain reliable results [34]. Particularly, a gradient descent optimization was performed for 1000 iterations. Furthermore, an early exaggeration trick was implemented when minimizing the KL-divergence. An additional momentum term of 0.5 was used for the first 250 iterations whereas it was increased to 0.8 for the rest. To perform t-SNE with Barnes-Hut approximation, we used the Rtsne package [35]. A comprehensive overview of parameters for deep stacked autoencoders and t-SNE are depicted in Table 3.

4.3 Demonstration and Evaluation

To demonstrate the contribution of our approach to AA, we examined four different substances that are currently being used in the production process of our industry partner, that are considered harmful for workers or the environment. These substances were selected by the ceramic manufacturer due to their individual risk profile: *Barium Carbonate, Instapak A, Sodium Borate,* and *Hydrogen Peroxide 30%.* We identified ten substitutes for every substance based on the proposed nearest neighbor search method (see fig. 3). The overall impression after an initial review of the ten suggestions by the expert was that for three of the four initial substances, at least one reasonable substitution candidate was provided through the machine learning process. Only in the case of *Instapak A*, substances cannot be easily exchanged since *Instapak A* is used as one component in a multi-component ready-to-use packaging foam provided by a third party. In other cases, a further investigation is necessary to back the positive first impression through professional literature and studies. An excerpt for the substance *Barium Carbonate* and the first five most important substitute candidates based on nearest neighbor search is listed in Table 4. We will explain the results of the evaluation in the case of *Barium Carbonate* in more detail. *Barium Carbonate* acts as a network converter in the glazing agent to increase the viscosity of the glaze mix in the fire kiln. It is also lowering the melting temperature of the glaze. At first sight, it is obvious that the first substitute candidate is related to the original substance based on its name. The base material of *Barium Carbonate* is *Barium*. The ML algorithm identified *Barium* as a substitute candidate for *Barium Carbonate* because the physio-chemical attributes of the two are almost identical. The second substitute, *Boron Nitride*, turned out to be impractical from the standpoint of technical product design. *Lead*, on the other hand, is a member of the SVHC list provided by ECHA and is even more dangerous than *Barium Carbonate*. Therefore, it was not considered in the further process. With *Carbonic Acid* we discovered a substance that fits the spectrum of substitute candidates from a technical, and safety standpoint. Thus, we proceeded with that substance to the next AA process step. What the evaluation of the first process step revealed is that the final choice of substitute candidates cannot be made by a data scientist or an algorithm alone. Instead, a domain expert from the manufacturing company must conduct a preliminary assessment to choose, which substitute candidates should be regarded in the further evaluation process. We also received the feedback that a function to exclude substances from the SVHC list from the suggestions would reduce the effort necessary for the manual review of the results.

Table 4. First five replacement substances for barium carbonate

Orig. substance	1st substitute	2nd substitute	3rd substitute	4th substitute	5th substitute
Barium carbonate CAS# 513-77-9	Barium CAS# 7440-39-3	Boron nitride CAS# 10043-11-5	Dicopper oxide CAS# 1317-39-1	Lead CAS# 7439-92-1	Carbonic acid CAS# 463-79-6

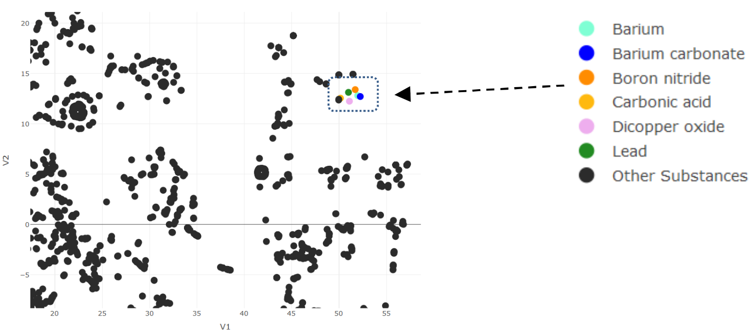


Figure 3. Visualization of substances with t-SNE (zoomed region)

The second step of our AA process was the comparison of the hazard and ecological factors that the new substances raise. Depending on the application context, a substitution should only take place, if the relevant risks for human and environment decrease due to the substitution. To compare the risks of both substances, we applied the GHS column model [18], which helps us to compare the two substances regarding the six criteria listed as columns in Figure 4.

	Barium carbonate CAS# 513-77-9					
	Acute health hazards (single exposure)	Chronic health hazards (repeated exposure)	Environmental hazards	Physico-chemical effects (fire, explosion, corrosion)	Hazards from release behaviour	Process-related hazards
	Acutely toxic substances/ mixtures, Cat. 4 (H302)	Safe substances on the basis of experience (e.g. water, paraffin and the like)	Substances/mixtures not hazardous to the aquatic environment (NWG, former WGK 0)	Non-combustible or only not at all readily flammable substances/mixtures (flash point of liquids > 100 °C, no H-phrase)	Non-dust-generating solids	Process index 0,25 according to TRGS 500
Risk potential:	2	0	0	0	0	0

Carbonic acid, zinc salt CAS# 51839-25-9					
Acute health hazards (single exposure)	Chronic health hazards (repeated exposure)	Environmental hazards	Physico-chemical effects (fire, explosion, corrosion)	Hazards from release behaviour	Process-related hazards
Safe substances on the basis of experience (e.g. water, paraffin and the like)		Substances/ mixtures not hazardous to the aquatic environment (NWG, former WGK 0)	Non-combustible or only not at all readily flammable substances/mixtures (flash point of liquids > 100 °C, no H-phrase)	Non-dust-generating solids	Process index 0,25 according to TRGS 500
0	0	0	0	0	0

Degree of Risk (numeric/colorized)

4	very high	very high	very high	very high	very high
3	high	high	high	high	high
2	medium	medium	medium	medium	medium
1	low	low	low	low	low
0	negligible	negligible	negligible	negligible	negligible

Figure 4. Hazard related comparison between substances

We compared *Barium Carbonate* and *Carbonic Acid*, which was ranked 5th on the substitute candidate list. As can be seen in Figure 4, *Barium Carbonate* and *Carbonic Acid* have an almost identical hazard profile. Acute health hazards that unfold at every exposure for *Carbonic Acid* pose only a low risk compared to a high risk of *Barium Carbonate*. That makes *Carbonic Acid* the substance of choice for manual tasks with substance exposure to the workforce.

5 Discussion of the Results and Conclusion

As ML is gaining momentum in many fields, new opportunities for solving information intensive tasks using ML techniques emerge. With alternatives assessment, we applied ML on an information intensive task. As emphasized before, the overload on information when it comes to assessing alternative substances along with the attitude to “never change a winning team” is what is causing many lost opportunities to transform production processes or recipes to more environmentally or workforce friendly variants.

The results of our research show that AA is a prime example for applying ML to ease existing problems, that are often not properly addressed due to lack of experts or due to resistance to change. Surveys reveal that *lack of information on alternatives* along with a *lack of relevant expertise and resources in companies* are two major inhibitors when it comes to AA [36]. We addressed the first issue by providing an IS and a database along with the identification of relevant data that fosters alternatives assessment. To approach the second problem, we proposed an ML technique that supports domain experts in finding reasonable substitution candidates that can be further assessed in terms of their economic aspects. The feedback that we received from our project partner so far was promising and the results are further evaluated by the product engineering department.

While the research project that is subject to this paper is still ongoing, future efforts shall focus on further aspects of AA. For example, on the integration of enterprise IS to improve the process of AA. Currently, each substance must be assessed manually in our application. Providing data of recipes directly from ERP systems would allow us to process large sets of data and find similar but less toxic substitute candidates on a broader scale. In addition, the economic aspect of a substitution will be focused on and added to our application.

6 Acknowledgment

The research described in this paper was supported by a grant from the German Ministry of Education and Research (BMBF), project name SiFa 4.0 (01IS17009B). Parts of the data was kindly provided by the European Chemicals Agency (ECHA).

References

1. Cusack P, Perrett T (2006) The EU RoHS Directive and its implications for the plastics industry. Plast Addit Compd 8:46–49
2. BizNGO (2013) The Commons Principles for Alternatives Assessment: Addressing Chemicals of Concern to Human Health or the Environment, https://www.turi.org/TURI_Publications/TURI_Chemical_Fact_Sheets/Commons_Principles_for_Alternatives_Assessment (Accessed: 19.07.2018)
3. Jacobs MM, Malloy TF, Tickner JA, Edwards S (2016) Alternatives assessment frameworks: Research needs for the informed substitution of hazardous chemicals. Environ Health Perspect 124:265–280
4. Geiser K, Tickner J, Edwards S, Rossi M (2015) The Architecture of Chemical Alternatives Assessment. Risk Anal 35:2152–2161. https://doi.org/10.1111/risa.12507
5. Rossi M, Tickner J, Geiser K (2006) Alternatives Assessment Framework of the Lowell Center for Sustainable Production. http://www.sustainableproduction.org/downloads/FinalAltsAssess06_000.pdf (Accessed: 21.08.2018)
6. Rossi M, Peele C, Thorpe B (2012) BizNGO Chemical Alternatives Assessment Protocol, https://www.bizngo.org/alternatives-assessment/chemical-alternatives-assessment-protocol (Accessed: 15.08.2018)
7. U.S. EPA (2011) Design for the Environment Program Alternatives Assessment Criteria for Hazard Evaluation, https://www.epa.gov/sites/production/files/2014-01/documents/aa_criteria_v2.pdf (Accessed: 25.08.2018)
8. Reihlen A (2011) Guide on Sustainable chemicals: decision tool for substance manufacturers, formulators and end users of chemicals, https://www.subsport.eu/wp-content/uploads/2011/10/UBA_guide_sustainalbe_chemicals.pdf (Accessed: 28.07.2018)
9. IC2 (2013) Alternatives Assessment Guide, http://theic2.org/article/download-pdf/file_name/IC2_AA_Guide_Version_1.1.pdf (Accessed: 30.07.2018)
10. Erickson B (2017) How many chemicals are in use today? Chem Eng News 95:23–24

11. Isaacs KK, Goldsmith MR, Egeghy P, et al (2016) Characterization and prediction of chem. functions and weight fractions in consumer products. Toxicol Reports 3:723–732
12. Hevner AR, March ST, Park J, Ram S (2004) Design Science - Hevner. Des Sci Inf Syst Res Author MIS Q 28:75–105
13. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A Design Science Research Methodology for Information Systems Research. J Manag Inf Syst 24:45–77
14. Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. MIS Q 37:337–355
15. Winter R (2008) Design science research in Europe. Eur J Inf Syst 17:470–475
16. Österle H, Otto B (2010) Consortium Research. Bus Inf Syst Eng 2:283–293
17. National Research Council (2014) A Framework to Guide Selection of Chemical Alternatives. Nat. Acad. Press, http://www.nap.edu/catalog/18872 (Accessed: 10.08.2018)
18. Smola T (2014) The GHS Column Model as an aid to selecting substitute substances, https://www.dguv.de/medien/ifa/en/pra/ghs_spaltenmodell/spaltenmodell_2017_en.pdf (Accessed: 12.07.2018)
19. Min R, Stanley DA, Yuan Z, et al (2009) A deep non-linear feature mapping for large-margin knn classification. Ninth IEEE Int. Conference on Data Mining, 2009. pp 357–366
20. Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans Pattern Anal Mach Intell 36:2227–2240
21. Keogh E, Chakrabarti K, Pazzani M, Mehrotra S (2001) Locally adaptive dimensionality reduction for indexing large time series databases. ACM Sigmod Rec 30:151–162
22. Salakhutdinov R, Hinton G (2007) Learning a nonlinear embedding by preserving class neighbourhood structure. In: Artificial Intelligence and Statistics. pp 412–419
23. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science (80-) 313:504–507
24. Larochelle H, Bengio Y, Louradour J, Lamblin P (2009) Exploring strategies for training deep neural networks. J Mach Learn Res 10:1–40
25. Davenport TH, Harris JG (2007) Competing on analytics : the new science of winning, 1st ed. Harvard Business School Press
26. Sanders NR, Manrodt KB (2003) The efficacy of using judgmental versus quantitative forecasting methods in practice. Omega 31:511–522
27. Halper F (2016) Operationalizing and Embedding Analytics for Action, https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/tdwi-operationalizing-embedding-analytics-for-action-108112.pdf (Accessed: 20.07.2018)
28. Gunning D (2017) Explainable Artificial Intelligence (XAI). DARPA
29. Gregor S, Benbasat I (1999) Explanations from intelligent systems: theoretical foundations and implications for practice. MIS Q 23:497–530
30. Lipton ZC (2016) The Mythos of Model Interpretability. ArXiv e-prints, June 2016.
31. Van der Maarten L., Hinton G (2008) Visualizing Data using t-SNE . Journ of Mach Learn Res 9 (2008) 2579-2605
32. Candel A, Parmar V, LeDell E, Arora A (2016) Deep learning with h2o. H2O Inc
33. Beygelzimer A, Kakadet S, Langford J, et al (2013) FNN: fast nearest neighbor search algorithms and applications. R Packag version 1
34. Wattenberg M, Viégas F, Johnson I (2016) How to use t-sne effectively. Distill 1:e2
35. Krijthe JH (2015) Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. R Packag version 013, URL https//github com/jkrijthe/Rtsne
36. Tickner J, Jacobs M (2016) Improving the identification, evaluation, adoption and development of safer alternatives: Needs and opportunities to enhance substitution efforts within the context of REACH. Environ Health Perspect. 2016 Mar;124(3):265-80

A Hierarchy of DSMLs in Support of Product Life-Cycle Assessment

Mario Nolte[1], Monika Kaczmarek-Heß[1], Andreas Fritsch[2], and Stefanie Betz[3]

[1] University of Duisburg-Essen, Essen, Germany
{mario.nolte,monika.kaczmarek}@uni-due.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
andreas.fritsch@kit.edu
[3] Furtwangen University, Furtwangen, Germany
besi@hs-furtwangen.de

Abstract. To support understanding and analysis of sustainability related aspects in organizations (e.g., via an assessment of a product's life-cycle from the cradle to the grave), various instruments, among others, in the field of conceptual modeling, have been proposed. Although existing tools and languages are, to some extent, indeed supporting the product Life-Cycle Assessment (LCA), our investigations show that a hierarchy of Domain-Specific Modeling Languages (DSMLs) is needed to satisfy advanced requirements. In this paper, as an innovation for the field of LCA, we propose an application of multi-level language architecture to design a hierarchy of DSMLs encompassing concepts for LCAs that can be detailed to specific industrial domains and local needs of enterprises. This enables a new generation of instruments allowing users to use and refine concepts, corresponding to their specific needs.

Keywords: LCA, sustainable development, multi-level modeling.

1 Introduction

To support the sustainable development (SD) of organizations, the awareness about ecological and social impacts of their products, potentially leading to unintended changes in the environment, needs to be increased. To increase that awareness, relevant information on potential impacts caused by all activities related to the production, usage and disposal of products, needs to be collected and used in decision-making processes [1, p. 226]. In this context *Life-Cycle Assessment* (LCA) has been established and used over the last decades to collect such information in a systematic manner [2]. Standardized by norms like ISO 14040 [3], LCA provides generic concepts and instructions that have been refined into several different assessment methods containing more specific concepts, which address the information needs within different industrial domains [4]. To support the application of LCA assessment methods various tools are provided, which produce complex results that are not easy to interpret and communicate, cf. [5]. In addition, currently existing LCA tools do not provide satisfactory support either [6,7], as results are not always transparent and traceable [8].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

To mitigate these challenges our earlier work shows how conceptual modeling can be used as an instrument to collect, structure, aggregate, and present data about potential ecological and social impacts of products along their entire life-cycle. For this, we proposed a modeling language *TracyML* [9] and a modeling method *ImpactM* [10]. Due to contestedness of the idea of SD [11] both methods do not provide a solution on their own, but allow to document relevant information required for the needs of the discursive decision-making like assumptions and information about system boundaries allowing for the comparison of results of different assessments.

Both languages are implemented in a *conventional language paradigm*, where the language specification is defined using a meta model that can be used to develop models one language-level below (i.e., on M_1). As both languages are based on ISO, therefore, for the sake of reuse, they are kept rather generic. As a result, although showing their applicability in different scenarios, those languages exhibit a similar deficiency as ISO 14040. Namely, to satisfy the specific information requirements of industrial domains or enterprises, first, a substantial effort needs to be invested to define and/or adjust the required concepts during the language use. Although it would be possible to propose a variety of Domain-Specific Modeling Languages (DSMLs) and various diagram types in LCA tools to avoid a need for such an adaptation, this approach would result in a threat to efficiency. Indeed, not only multiple DSMLs and diagram types would have to be developed and maintained, but also relevant information, e.g., on typical impacts related to domain-specific resources, would have to be provided during the language use, which results in considerable time and cost expenditures.

To overcome these challenges the goal of our research is to propose a hierarchy of DSMLs spanning through: (1) a *reference Domain-Specific Modeling Language* (rDSML) [12, p. 321], which includes generic concepts for LCA, that can be refined to (2) *specific industrial domains*, up to (3) languages with a high level of specificity for certain enterprises as *Enterprise-Specific Modeling Languages* (ESMLs), cf. [13]. Our proposal addresses advanced requirements, as there is a need to provide software support for different user groups and purposes. On the one hand, researchers need generic and flexible tools, which allow for modelling of standard and diverging scenarios. On the other hand, for industry users efficiency and effectiveness is most important. They need pre-specified and easy-to-use software that can be easily parameterized [14]. At the same time, a major challenge in LCA and weakness of available LCA software is ensuring comparability and compatibility of user models [6]. Within the concept presented in this paper, the industry- and enterprise-specific modeling languages form the basis for efficient and easy-to-use software tools that cater for the specific requirements of enterprises and industries. Hereby, the underlying integrated hierarchy provides the conceptual foundation for the specific modeling languages, and ensures the comparability and compatibility of user models.

This contribution follows the design-oriented research paradigm [15]. The resulting IT artifact (i.e., the hierarchy of DSMLs developed in an iterative manner) aims at providing a benefit to organizations by supporting LCA and addressing the above-mentioned challenges. To create the targeted hierarchy, we follow the language design method proposed by [16], which provides a macro-process and corresponding roles, as well as a set of guidelines. In this paper we focus on three selected outcomes, i.e., clarification of goals and scope of the solution, identified requirements, and a resulting hierarchy of DSMLs. We also present a realistic scenario to illustrate our vision.

The paper is structured as follows. First, a short overview on general ideas of product LCA is provided. Then, an exemplary scenario follows, which is used to explain the vision of our research. Next, requirements towards a language architecture are shortly discussed. Then, we present the hierarchy of DSMLs in form of a multi-level model.

2 (Product) Life-Cycle Assessment

LCA studies aim at identifying all relevant impacts through-out the life-cycle of a product. Thereby, the *product system* is defined as the life-cycle containing different states of a product, from extraction of the necessary raw materials, via production and assembly of its components to usage and final disposal or recycling [3]. The ISO 14040 standard lays out basic requirements to avoid biased studies and inappropriate claims [17]. Thus, it provides a common language and guidelines how to apply it. The four main steps for conducting an LCA study comprise [3]: (1) goal and scope definition, (2) life-cycle inventory analysis, (3) life-cycle impact assessment, and (4) interpretation. During the *goal and scope definition*, in particular the system boundary is to be defined. Depending on information needs and practical constraints, one may, e.g., choose to address the whole life-cycle from resource extraction to disposal ("cradle to grave"), or just the assembly of some parts ("gate to gate") and define cut-off criteria (e.g., excluding material below a specific weight). The second phase, the *life-cycle inventory analysis*, is where the data collection and definition of indicators takes place. Impacts are modeled in the third phase, *life-cycle impact assessment*. This is typically done using the concepts *impact category* and *category endpoint*, cf. Tab. 1. Note that these are very broad concepts and, as *different assessment methods* are available, cf. [18], their usage differs significantly. In the following, we use the term *impact* to describe an effect that may happen on a global (e.g., climate change), regional (e.g., smog, eutrophication) or local level (e.g., acidification of water), and can be traced back to a product system. Correspondingly, an *endpoint* may represent an entity that is affected by an impact (e.g., freshwater supply in a lake). Finally, in the last step, the assessment results are interpreted and communicated.

Table 1. Selected concepts proposed in ISO 14040 for LCA [3, pp. 7-14]

Term	*Definition*
product	"any goods or service"
product system	"collection of unit processes with elementary and product flows, performing one or more defined functions, and which models the life cycle of a product"
sys. boundary	"set of criteria specifying which unit processes are part of a product system"
funct. unit	"quantified perform. of a product system for use as a reference unit"
impact category	"class representing environmental issues of concern to which life cycle inventory analysis results may be assigned"
category endpoint	"attribute or aspect of natural environment, human health, or resources, identifying an environmental issue giving cause for concern"

Due to the applicability of the standard to various domains and contexts, the *concepts are specified on a general level* and the standard itself offers freedom to users in defining the functional unit, system boundary, data sources etc. As a result, a large amount of literature, standardization processes [19] and databases exists, which reference the ISO-concepts, e.g. [4]. Additionally, there are efforts to apply LCA to the social dimension of sustainability [20], and advance the method to provide an integrated ecological, social and economic perspective [21].

3 Motivating Scenario

Consider two companies intending to improve their sustainable development: Company T producing T-Shirts out of natural fiber, and Company S producing shelves composed out of wood and metal. In particular, those companies wish to reduce their ecological and social impact, while ensuring on-going and long-term economic success. The comprehensiveness of the LCA-approach as the "gold standard" of sustainability assessment has already convinced them to follow a life-cycle perspective on their products. They expect that this will help them identify all relevant impacts and avoid problem shifting between them.

However, once our companies start the analysis, it becomes unclear to them (as in-depth LCA expertise within companies is missing), what potential impacts would be relevant for the analysis in question. Likewise, both companies face difficulties with decisions to be made, e.g., what phases actually to include in the assessment. Thus, both companies are challenged with a need to reconstruct their domain knowledge and information needs using the generic concepts of the ISO-standard. As a result, due to practical constraints like data availability [19], and also different information interests, the resulting models created by companies are on various levels of abstraction and use data on various levels of detail.

Next, the *availability of resources* for raw materials on the local and world-wide level needs to be considered. While the growth-rate of different kinds of natural fiber is of interest for Company T, Company S has interest in the growth-rate of different kinds of wood. Because the shelves contain also metal parts, Company S is also interested in the availability of metal stocks (e.g., Bauxite as ore for aluminum), which can be expressed through the expected date of depletion of specific mines on a local level, or through resources expected in the earth crusts and anthropogenic stocks (e.g., used in cars) on a global level. Regarding the *raw materials* both companies share the interest in economic data, like sales and average prices, and the recycling code that might help at a later point in time determine the potentials for recycling raw materials.

Coming to discuss relevant potential and actual social and ecological impacts that can be traced back to raw materials, it becomes clear that a huge variety of impacts and related endpoints can be calculated by different assessment methods and related indicators. Indeed, regard, e.g., social impacts: while for the assessment of wood-based production systems impacts related to health & safety, employment or equal opportunities of forestry workers might be relevant, for fiber-based products impacts like toxic emissions related to the dying of cotton is of importance. Here it can be differentiated between effects from inhalation, which can get chronic or carcinogenic, and effects stemming from ingestion. Furthermore, because Company T intends to avoid the consumption of natural fiber as a raw material having a high need for water consumption with respect to its origin, this company defines the availability of water as an *endpoint.* Therefore, it requires information about the average amount of water for the raw material of natural fiber in general, as well as the amount of water for natural fiber consumed in a specific farm. Also, the soil of a farm is in this context important and treated as an endpoint, which is characterized by its fertility and erosion. In the long-term, Company T aims to lower the rate of farms supplying natural fiber, which have too high water consumption or a high grade of erosion. In turn, Company S intends to make use only of wood that stems from a forest management targeting at the production of wood and not from forests managed for other purposes, e.g., conservation, recreation. This information needs to be modeled as endpoints. In addition, since this company also intends to provide financial support for non-production forests that are untouched by human influence, the grade of hemeroby (e.g., ahemerobic for non-influenced forests) should be included in models too.

Finally, it is also important for both companies to have specific quality-related information on raw material used. Here, it is assumed that *branchiness* and the *durability class* in accordance to DIN EN 350-2 [22] are considered as relevant for the raw material of wooden products for Company S. In turn, Company T is interested in codes provided in the DIN 60001-1 [23] for textile materials, as well as in *elasticity* of different kinds of fiber, and in *tensile strength* assessing the capacity of fiber or other materials to elongate without tearing apart.

Table 2. Aspects of interest and related LCA concepts - overview

	Company T	*Company S*
Product	T-Shirts	Shelves
Raw material	Natural fiber (avg./sales price; recycling code)	Wood (avg. and sales price; recycling code); metal (avg. and sales price; recycling code)
Endpoint	Soil of farms (location; fertility; erosion)	Forest stands (location; commercial forest); mines (location; downhole)
Impact	Resource depletion natural fiber (kind; renewal rate; consump. worldw.; stock worldw.); virtual water	Resource depletion wood (kind of fiber; renewal rate; consumption worldw); resource depletion metals (kind; consumption worldw.; sum known resources); pot. social and environ. impacts
Resource	Stocks of different kinds of natural fiber, e.g., cotton, linen (global and local)	Stocks different kinds of wood, e.g. birch, spruce (global and local); stocks of different kinds of metal, e.g. aluminum
Level of detail	Typical / average information; specific stocks of resources	Typical/avg. inform.; resources specific stocks; individual product dependent information for certificates demanded by customers
Quality interests	Elasticity; tensile strength, Code DIN 60001-1	Branchiness (high, average, low); durability class ('1'=very durable, '5'=non-durable)

Tab. 2 gives a structured overview of the scenario. In the first three lines concepts from ISO 14040 are assigned to the specific interests of both companies. The fourth row presents different kinds of resources. Terms in the brackets should give an impression on possible exemplary attributes, values or further concepts that are necessary to support the impact assessment as sketched above. The two last lines indicate individual quality-related interests and levels of details that should be satisfied by conceptual models that support this scenario. Please note that only the ISO concepts themselves are readily available and all other ones need to be modeled or reconstructed, which significantly hampers the productivity of the analysis. Finally, due to, e.g., different modeling decisions and data/concepts used, comparing underlying models and obtained results would be challenging.

4 Vision and High-Level Goals

Our aim is to offer a hierarchy of DSMLs encompassing a *reference DSML* (rDSML) including concepts for conducting LCA, which are later refined to specific industrial domains with an increasing level of specificity, up to certain enterprises as *Enterprise-Specific Modeling Languages*, cf. Fig. 1. Thus, within the offered hierarchy, both companies would have an access to a domain-specific language for conducting LCAs in their industry, suited to their needs, supporting them in conducting the desired assessment, and ensuring required comparability and transparency of achieved results.

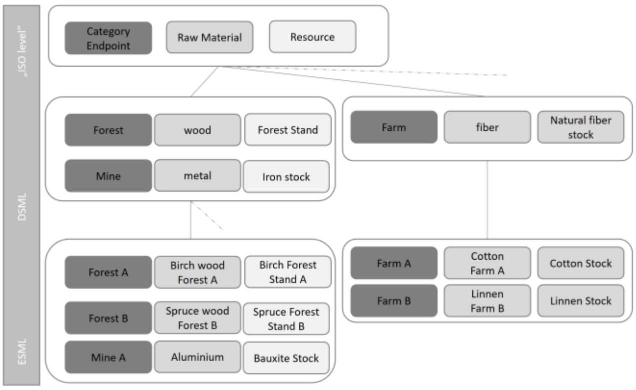


Figure 1. Vision: A hierarchy of DSMLs

And so, while the rDSML provides generic ISO concepts, the refined DSMLs provide specific, semantically rich concepts with a wide range of properties appropriate to describe the corresponding domain. These concepts store relevant information for the needs of assessment, which supports its productivity. Users of the language, even if they are not experts in LCA, can benefit from the incorporated (domain-specific) knowledge on how to conduct the desired analysis. This knowledge encompasses, among others, a set of impacts, their indicators, assessment methods, requirements regarding the data to be used, as well as, if possible, the required data itself.

The hierarchy allows both companies to conduct the analysis on the local and global scale on different levels. This means that they can use one of the more specialized

DSMLs to conduct the analysis of interest, and then aggregate (“bottom-up”) the results by moving up along the hierarchy. It is also possible to “drill down” in the model to individual localized impacts, that are relevant to specific stakeholder groups, e.g., to identify impacts on a specific ecosystem caused by water extraction from a lake (environmental), or to identify social problems like excessive working hours at a specific site. This would enable stakeholders’ engagement and provide the possibility to identify and address concrete problems. Finally, guidelines (e.g., for interpretation and presentation) are embedded into the language, enriched by a visual notation following cognitive principles of information design to facilitate communication and understanding of the process and its results.

In order to: (1) allow for application of concepts from different *impact assessment methods* in tandem, (2) allow users to access all classification (specificity) levels they are interested in, as well as (3) ensure the comparability of achieved results (e.g., by models explicating assumptions and information about the system boundaries, cf. [10]), all of the DSMLs in the hierarchy are integrated. This integration is achieved through the application of the same language architecture, through the refinement of concepts from the reference DSML (vertical integration), and also through the definition of aligning horizontal relationships between concepts. In consequence, the hierarchy of DSMLs offers the required transparency for evaluating and comparing the achieved assessment of product systems, cf. [6]. In addition, the presented hierarchy is adaptable, meaning that once new developments (e.g., new *impact assessment methods*) are known, they may be accounted for within the appropriate DSML.

Currently, there exists a wide range of LCA tools: generic expert tools and specialized ones focusing on specific areas [24]. Considering that the conflict of standardization versus extensibility is one of the major challenges discussed in LCA literature [5,6,14], this hierarchy of DSMLs should allow for a new generation of LCA tools enabling users to build semantically rich models while reusing existing concepts. In particular, the resulting tools should have flexible architecture allowing for integration of standards and best practices (on higher levels), while providing extensibility to account for different application scenarios and methodological advancements (on lower levels).

To summarize, the hierarchy of DSMLs is to support the following high-level goals: (G1) Provide a support for a wide range of different perspectives prospective users may be interested in, by providing a hierarchy of vertically and horizontally integrated DSMLs; (G2) Offer semantically rich concepts and required information to support the assessment process; (G3) Account for existing standards and branch-specific methods in a way to ensure comparability of achieved results, (G4) Support conducting analysis on a local and global level; (G5) Support extensibility and adaptation to account for the relevant changes; and (G6) Support both the productivity and reuse of the approach.

5 Requirements

In line with the method followed [16], based on the identified goals (Sect. 4), a set of identified application scenarios, analysis of LCA methods and instruments, as well as problems reported in the literature, e.g., [5,6,14], we have defined a set of requirements towards (1) the scope of the targeted solution (i.e., concepts and functionalities required to conduct LCA in different domains), and towards (2) the language architecture of the targeted artifact. Due to the space restriction, we discuss a few selected requirements towards language architecture only.

R1: Accounting for a hierarchy of professional terminology encompassing various classification levels. *Rationale*: Our goal is to provide support for a wide range of different perspectives of prospective users (cf. G1). Considering that different users consider involved concepts at different levels of specificity, there is a need to account for a hierarchy of professional terminology. For instance, the LCA concepts such as endpoint and raw material can be of interest on different levels of detail, e.g., as a raw material as such, as Plate of Wood, as Plate of Birch Wood only, or as instances of the latter, i.e., some specific forests. Therefore, the language should provide information on relevant aspects accounting for generic terms, their types, categories and instances. *Discussion:* In the current practice DSMLs are built by defining a meta model using some general-purpose modeling language, e.g., Meta Object Facility in UML [12]. This meta model describes concepts that users may use to create models. Thus, a modeling language usually encompasses two layers: specification (i.e., the definition of a DSML) and language application, cf. [25]. However, in our case, the language itself spans an arbitrary number of classification levels, not only two. Therefore, instead of emulating several meta-levels within two levels, or using artificial workarounds [26], a more natural way to define the desired hierarchy is the use of multi-level modeling [27]. Multi-level modeling refers to a language architecture that allows for an arbitrary number of classification levels being represented within a single body of model [27].

R2: Facilitating horizontal and vertical integration. *Rationale*: A rapid growth of various initiatives both horizontally (e.g., ecological and social aspects) and vertically (e.g., specific methods for some types of resources/some industries) causes a need for integration as well as ensuring comparability between achieved results (cf. G1 and G3). *Discussion:* In order to support integration (and also avoid redundancy), language users should be able to state what they know as soon as they know it, cf. [28]. For instance, already on the level of a rDSML encompassing, e.g., a concept *NaturalResource*, we would like to state that on the instance level (which is a few classification levels below) an attribute summing up production per year will be applicable. To be able to define it however, we would need a deep instantiation mechanism [27] that allows us to define some properties on a higher level and defer their instantiation to some not directly proceeding classification level. Such a mechanism is offered by multi-level modeling approaches and is not supported by traditional ones.

R3: Providing support for productivity of modeling and reuse. *Rationale*: The scenario indicates that a wide range of aspects of a concept (e.g., “resource”) should be accounted for in an integrated manner. Furthermore, their description should be rich enough to support various analyses regarding, e.g., their general character (e.g., hazardous) or specific characteristics (e.g., the rate of growth, societal relevance) [5, 19]. To this aim, we require domain-specific concepts with a rich set of attributes [29]. This is to support modeling productivity, i.e., productivity of creating, analyzing and modifying models (i.e., the time needed to accomplish those tasks [30]). However, on the other hand, at the same time we want to provide a set of generic concepts, so that they may be reused in a wide range of (not yet known) scenarios, cf. G6. *Discussion:* If we are modeling using the conventional approach and are limited to two classification levels, we need to face a well-known conflict between the level of reuse and modeling productivity [12]: the more specific a concept is, the lower range of reuse there is. On the other hand, the more general the concept is, the wider is the range of reuse, but also the lower is the productivity of modeling. As already discussed by other authors, e.g., [12], a satisfactory solution to this conflict comes with a multi-level modeling language architecture where we can account for both generic and specific concepts at the same time through multiple classification levels and, on demand, select those that suit our modeling needs.

R4: Incorporating relevant knowledge within the language. *Rationale*: Conducting a life-cycle assessment requires, among others, (1) information on potential impacts, their indicators as well as reference values; and (2) a level of expertise and experience that may not be available in many organizations [5, 19]. *Discussion:* Incorporating relevant knowledge within the language specification implies, among others, assigning states to (meta) classes (e.g., stating that the value of recycling code for Plate of Wood is 'FOR'). This however, is impossible in the traditional language architecture due to the iron rule of the type/instance dichotomy [12]. Therefore, we would need to always provide this information on the level of language application only. Considering it, application of multi-level modeling seems reasonable, as it offers "clabjects" [27], i.e., concepts having the characteristics of classes (defining a structure) and objects (having a state).

R5: Equipping models and their elements with behavior. *Rationale*: (Meta) classes (e.g., *NaturalResource*) have features that need to be derived or calculated based either on the content of the model (e.g., calculation of an impact profile), or based on the data acquired from external sources [7] (e.g., to obtain some environmental data for the needs of the assessment calculation), cf. G2. In addition, we would like to execute operations (automated analysis) on model elements not only on the level of objects, but also on the level of classes (e.g., calculation of global and local impact on the forest, cf. G4). *Discussion:* While it is possible to define language concepts having attributes constituting abstractions over desired data, the problems with the (automated) acquisition/calculation of those values emerge. The reason for this is not necessarily connected with the modeling paradigm as such, but rather with the programming languages used to create modeling tools. Modeling tools are usually developed using mainstream object-oriented programming languages, which feature only one classification level. Thus, types or even meta types are represented as objects by overloading the M_0 level of a programming language [28]. Therefore, a common representation of code and model is not possible and a model-code synchronization is required [28]. Thus, not only a recompilation step of modeling tools is required whenever we want to change something in the language specification (cf. R6), but also equipping model elements with operations (e.g., allowing to retrieve actual state of the environment) is hardly conceivable [28,31]. Therefore, a satisfactory solution seems to be the application of an integrated modeling and programming, i.e., using a modeling approach coming with a language execution engine.

R6: Ensuring extensibility and adaptability of the hierarchy without losing a corresponding tool support. *Rationale*: As already mentioned, there is a need to account for new developments and additional knowledge (G5), e.g., on impact caused by different substances. Therefore, the proposed hierarchy should be easily extensible and adaptable [19]. As an appropriate tool support is necessary to conduct the assessment process, those changes should not lead to losing this support. *Discussion:* If we decide to use a conventional approach and a tool based on the semantics of dominant object-oriented programming languages, even if users would have an access to language specification and would be able to extend it and adapt it to their needs, a recompilation step would be required to account for the changes in the corresponding tool [32]. A satisfactory solution to this problem again comes with the application of the multi-level approach, where the border between language specification and application is blurred, implying that by changing a multi-level model, a user may adapt the language to his/her needs [30]. If we use the integrated modeling and programming environment, also the tool-support will not be lost, cf. [28].

6 A Hierarchy of DSMLs for LCA

As discussed in the previous section, when it comes to the application of conventional two-level modeling paradigm, although its application is *technically* possible, it imposes limitations, which hinder us from delivering a satisfactory language specification, i.e., a solution without workarounds, overloaded levels, model redundancy and accidental complexity [12, 26, 33]. Therefore, we turn our attention to multi-level modeling with an integrated programming environment to propose a hierarchy of DSMLs.

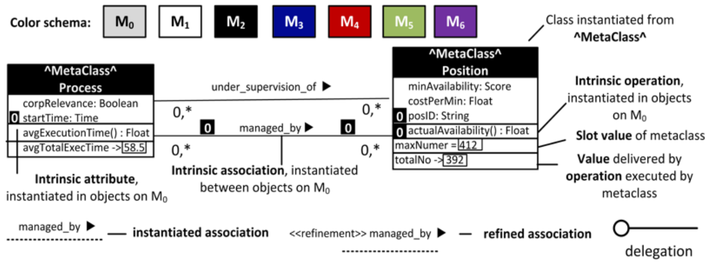


Figure 2. Concrete syntax of FMML^X, based on [12]

Multi-level modeling refers to a language architecture that allows for an arbitrary number of classification levels which are represented within a single body of model content, cf. [27]. As there is no strict division between language specification and language application, all languages (generic, regional, local ones) are integrated in one language architecture and thus, users can access all classification levels they are interested in. We can also benefit from such features as relaxed type/instance dichotomy, deferred instantiation, or defining level-crossing relations, cf. [31]. To illustrate the prospects of multi-level modeling, let us consider an excerpt of a multi-level model of resources created using a Flexible Meta Modeling and Execution Engine ($FMML^x$) [12]. We have selected $FMML^x$ for our design due to the fact that, to the best of our knowledge, $FMML^x$ as the only one has an integrated language execution engine offered by a supporting tool XModeler [12]. This feature allows us, among others, to equip models with behavior and provide support for computational analysis. Whereas the detailed description of $FMML^x$ may be found in [12, 28], Fig. 2 presents its concrete syntax. Apart from the "traditional" modeling constructs such as classes, attributes, operations and relationships, it is possible to defer an instantiation of all modeling constructs by assigning them so called level of *intrinsicness*, which tells at which level of classification a given property will be instantiated.

Fig. 3 presents an excerpt of the designed hierarchy of DSMLs in the form of a multi-level model. Please note that for readability purposes, only selected concepts, selected attributes and operations derived from our scenario assigned to different levels of classification are presented. By supporting multiple classification levels, $FMML^x$ offers the possibility to define and use concepts that correspond directly with the desired level of details (R1). Thus, we have a possibility to account for the fact that such concepts as *NaturalResource* or *RawMaterial* span multiple levels of classifications with categories, types and instances, cf. ① Fig. 3, without the need for overloading some level or applying some other workaround. At each level of classification, we have the possibility to express relevant information, making the model semantically rich (thus, we support the productivity of modeling as well as enable various analyses), and at the same time facilitate its reuse (R3). Regarding the latter, consider, e.g., attributes, operations and relationships defined for the concept *NaturalResource* (M_3), which reflect the domain knowledge derived out of the scenario on this classification level about this concept. Please note that a majority of those characteristics will be instantiated (i.e., assigned with values) only a few classification levels below (cf. the assigned level of intrinsicness). For example, the attribute *sum production per year local* (sumProdPyL) is instantiated in the local stocks of natural resources, be they stocks of cotton residing on farms or stocks of caolinite (mineral used for aluminum) residing on mines ②.

Now, while we move along the created hierarchy ③ (e.g., the chain starting from *RawMaterial* (M_4) via *PlateOfWood* (M_3) up to a *specific Plate of Wood* (M_2) and its instances), on the one hand, we instantiate the concepts, i.e., the relevant attributes are assigned with values (e.g., ④ the recycling code which is the same for all wooden products, and which is assigned on level M_3, cf. R4) and relevant operations aggregating, calculating or acquiring data from external sources may be executed (e.g., cf. ⑤, *calculateConsumptionWoW*(), or *sumAvailableResources*() defined for *NaturalResource*, M_4, cf. R5). On the other hand, we specialize those concepts, i.e., additional attributes, operations and relationships may be added to make concepts more and more specific (cf. ⑥, additional attributes defined for the concept *AbioticResource*, M_2). Furthermore, not only attributes are getting refined while moving along the hierarchy. This applies also to defined relationships. For instance, while we define that a *RawMaterial* (M_4) *stems from* some *NaturalResource* (M_3) and defer instantiation of this relationship to M_0, on the level of *CottonSheet* (M_2) we refine this relationship, and state that it can only stem from concepts instantiated from *CottonStock* (M_1) ⑦.

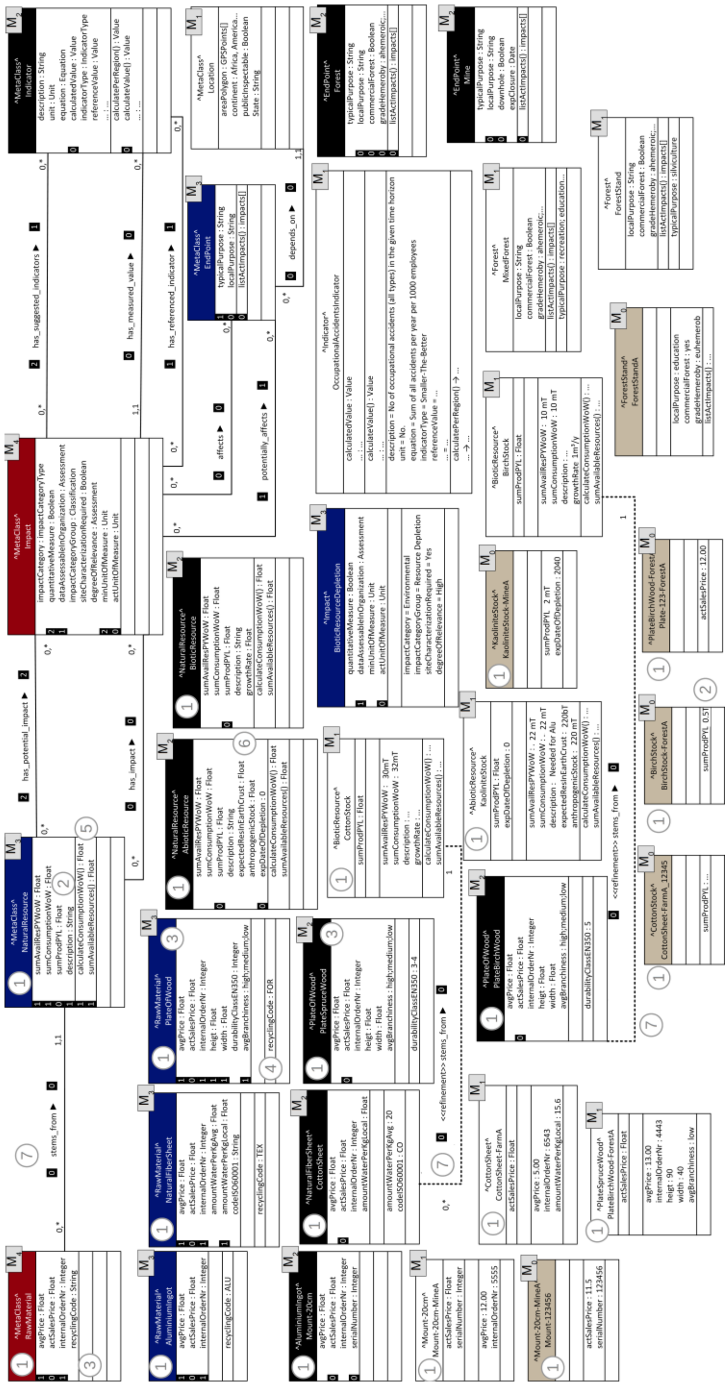


Figure 3. An excerpt of the designed hierarchy of DSMLs

The designed hierarchy fulfills the requirements (cf. Sect. 5) as summarized in Tab. 3. However, please note that the presented multi-level model may seem to be too complex for users [34]. The complex representation serves to exemplify how users can access concepts underlying their DSML and thus, benefit from greater transparency.

Table 3. Summary of evaluation against the requirements

Req.	*Evaluation and scenario-dependent solution*
R1	Concepts relevant for SD can be refined to domain specific concepts (e.g., Raw Material to Natural Fiber Sheet) up to enterprise specific concepts (types of Cotton Sheet) and instances.
R2	Domain-dependent attributes like recycling codes for Raw Material can be set once. Derived concepts on deeper levels are related to this information, but can be extended, as the vertical integration of Raw Material shows.
R3	By offering concepts on different classification levels we support both productivity and reuse at the same time (e.g., we offer both an abstract concept Resource as well as a set of its specific types and instances).
R4	Thanks to relaxed type/instance dichotomy we may assign state to classes, and thus, e.g., state what is the recycling code for a Plate of Wood.
R5	As in FMML^x a class is an object [12], operations can be not only specified for classes but also executed on them (e.g., calculateConsumptionWoW()).
R6	Thanks to a common representation of model and code provided by XModeler [12], a multi-level model may be extended without a need for recompilation.

7 Conclusion

In this paper, we argue that in order to support the LCA of products in a satisfactory manner, there is a need to design a hierarchy of DSMLs. To this aim we presented in this paper the clarification of the goals and scope of such a hierarchy, the requirements towards the language architecture as well as an excerpt of the current state of the resulting hierarchy of DSMLs. We also pointed that the results of our work can be used for building a new generation of LCA tools.

A few important limitations of our work need to be mentioned. Firstly, although the results of our work applied to different scenarios seem to be promising, they also point to the need for further extensions. Namely, additional work is required to incorporate further domain-specific LCA methods and techniques, and ensure integration to existing databases. In addition, as our ultimate goal is to support analysis targeting strong sustainability, or supporting at least a substitution strategy, cf. [10], there is a need to further extend the ecological, economic, and social aspects accounted for.

Secondly, additional work is required to reduce the complexity of a resulting multi-level model. While the complexity of multi-level models is not a novel phenomenon, cf. [31, 34], it should be addressed since it makes the models difficult to interpret, thus potentially inhibiting the added analysis capabilities we aim at. Here the integrated modeling and programming environment XModeler [12] comes into play with the promise of being able to define various perspectives adjusted to the information needs of prospective users. Finally, it has to be remarked, that further limitations need to be accounted for which are inherent to LCA itself. They comprise boundary issues as described by Hovorka et al. [35], uncertainty regarding ecological causes and effects [36], as well as the appropriateness of a systems' perspective for social aspects [37].

References

1. Melde, T.: Nachhaltige Entwicklung durch Semantik, Governance und Management. Springer, Wiesbaden (2012)
2. Finkbeiner, M., Inaba, A., Tan, R., Christiansen, K., Klüppel, H.J.: The new international standards for life cycle assessment: ISO 14040 and 14044. Int J of LCA 11(2), 80-85 (2006)
3. ISO: DIN EN ISO 14044 - Umweltmanagement - Ökobilanz - Anforderungen und Anleitungen (2006)
4. Bach, V., Berger, M., Forin, S., Finkbeiner, M.: Comprehensive approach for evaluating different resource types - case study of abiotic and biotic resource use assessment methodologies. Ecological Indicators 87, pp. 314-322 (2018)
5. Hellweg, S., Milà i Canals, L.: Emerging approaches, challenges and opportunities in life cycle assessment. Science 344 (6188), 1109-1113 (2014)
6. Kuczenski, B., Marvuglia, A., Astudillo, M.F., Ingwersen, W.W., Satterfield, M.B., et al.: LCA capability roadmap-product system model description and revision. Int J of LCA 23(8), 1685-1692 (2018)
7. Kalverkamp, M., Karbe, N.: Comparability of life cycle assessments: Modelling and analyzing LCA using different databases. In: Pehlken, A., Kalverkamp, M., Wittstock, R., eds.: Cascade Use in Technologies 2018, pp. 51-63. Springer (2018)
8. Gardner, T.A., Benzie, M., Börner, J., Dawkins, E., Fick, S., Garrett, R., Godar, J., Grimard, A., Lake, S., Larsen, R.K., et al.: Transparency and sustainability in global commodity supply chains. World Development (2018)
9. Betz, S., Fritsch, A., Oberweis, A.: TracyML - a modeling language for social impacts of product life cycles. In Cabanillas, C., Espana, A., Farshidi, S., eds.: Proceedings of the ER Forum 2017 and the ER 2017 Demo track (2017)
10. Nolte, M., Kaczmarek-Heß, M.: Product life-cycle assessment in the realm of enterprise modeling. In: The Practice of Enterprise Modeling 2017, pp. 187-202 (2017)
11. Jacobs, M.: Sustainable Development as a Contested Concept. In: Dob, A., ed.: Fairness and Futurity: Essays on Environmental Sustainability and Social Justice, pp. 21-45. Oxford University Press, Oxford (1999)
12. Frank, U.: Multilevel modeling - toward a new paradigm of conceptual modeling and information systems design. BISE 6(6), pp. 319-337 (2014)
13. Becker, J., Janiesch, C., Seidel, S., Brelage, C.: A framework for situational and evolutionary language adaptation in information systems development. In: Advances in Information Systems Development, pp. 103-113. Springer (2007)
14. Unger, N., Beigl, P., Wassermann, G.: General requirements for LCA software tools. Institute of Waste Management, BOKU–University of Natural Resources and Applied Life Sciences, Vienna Austria (2004)
15. Österle, H., et al.: Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. Zeitschrift für betriebswirtschaftliche Forschung 62(6), 664-672 (2010)
16. Frank, U.: Outline of a Method for Designing Domain-Specific Modelling Languages. ICB Research Report 42, University of Duisburg-Essen, Essen (2010)
17. Finkbeiner, M.: The International Standards as the Constitution of Life Cycle Assessment: The ISO 14040 Series and its Offspring, pp. 85-106. Springer (2014)

18. Bare, J.C., Hofstetter, P., Pennington, D.W., De Haes, H.A.U.: Midpoints versus endpoints: the sacrifices and benefits. Int J Life Cycle Assess 5(6), 319 (2000)
19. Finkbeiner, M., Ackermann, R., Bach, V., Berger, M., Brankatschk, G., Chang, Y.J., et al.: Challenges in Life Cycle Assessment: An Overview of Current Gaps and Research Needs, pp. 207-258. Springer (2014)
20. Andrews, E.S., Barthel, L.P., Beck, T., Benoît, C., Ciroth, A., Cucuzzella, C., Gensch, C.O., et al.: Guidelines for Social Life Cycle Assessment of Products. United Nations Environment Programme (2009)
21. Onat, N.C., Kucukvar, M., Halog, A., Cloutier, S.: Systems thinking for life cycle sustainability assessment: a review of recent developments, applications, and future perspectives. Sustainability 9(5), 706 (2017)
22. DIN, E.: 350:2016-12. Durability of wood and wood-based products - Testing and classification of the durability to biological agents of wood and wood-based materials. Deutsches Institut for Normung (2016)
23. DIN: 6000-1:2001-05. Textile Faserstoffe - Teil 1: Naturfasern und Kurzzeichen. Deutsches Institut for Normung (2001)
24. Hitchcock, D., Schenk, R., Gordy, T.: Directory of Sustainability Life Cycle Assessment Tools. ISSP (2011)
25. de Lara, J., Guerra, E.: Domain-specific textual meta-modelling languages for model driven engineering. In: Vallecillo A., Tolvanen JP., Kindler E., Störrle H., Kolovos D. (eds) Modelling Foundations and Applications. ECMFA, pp. 259-274. Springer, Berlin (2012)
26. Atkinson, C., Kühne, T.: Reducing accidental complexity in domain models. SoSyM 7(3), 345-359 (2008)
27. Atkinson, C., Kühne, T.: The essence of multilevel metamodeling. In: Proc. of the 4th Int. Conf. on The Unified Modeling Language, Modeling Languages, Concepts, and Tools, pp. 19-33. Springer, London (2001)
28. Frank, U.: Designing models and systems to support IT management: A case for multilevel modeling. In: Atkinson, C., Grossmann, G., Clark, T., eds.: MULTI@MoDELS, ceur-ws.org, pp. 3-24 (2016)
29. Frank, U.: Multi-perspective enterprise modeling: foundational concepts, prospects and future research challenges. SoSyM 13(3), 941-962 (2014)
30. Frank, U.: Power-modelling: Toward a more versatile approach to creating and using conceptual models. In: BMSD'14, pp. 9-19 (2014)
31. Kaczmarek-Heß, M., de Kinderen, S.: A Multilevel Model of IT Platforms for the Needs of Enterprise IT Landscape Analyses. BISE 5 (2017)
32. Atkinson, C., Gerbig, R., Fritzsche, M.: A multi-level approach to modeling language extension in the enterprise systems domain. IS 54, 289-307 (2015)
33. Lara, J.D., Guerra, E., Cuadrado, J.S.: When and how to use multilevel modelling. ACM Trans. Softw. Eng. Methodol. 24(2), 12:1-12:46 (December 2014)
34. Kaczmarek-Heß, M., Nolte, M., Fritsch, A., Betz, S.: Practical experiences with multi-level modeling using FMML^x. In: 5th intl. Works. on Multi-Level Modelling (2018)
35. Hovorka, D.S., Labajo, E., Auerbach, N.: Information systems in environmental sustainability: Of cannibals and forks. In vom Brocke, J., Seidel, S., Recker, J., eds.: Green Business Process Management, pp. 59-72. Springer, Berlin (2012)
36. Adensam, H., Ganglberger, E., Gupfinger, H., Wenisch, A.: Wieviel Umwelt braucht ein Produkt? Österreichisches Ökologie Institut, Wien (2000)
37. Schultze, U.: What kind of world do we want to help make with our theories? Information and Organization 27(1) 60 - 66 (2017)

A Survey of Smart Energy Services for Private Households

Ute Paukstadt

University of Münster, European Research Center for Information Systems (ERCIS), Münster, Germany
ute.paukstadt@ercis.uni-muenster.de

Abstract. The energy sector is challenged by the ongoing digitalization with emerging smart energy products and services. Smart energy products such as smart meters leverage innovative smart energy services promising both new business opportunities and values for customers. Smart products and services could enhance energy efficiency as well as enable private households to produce their own energy. Although services are regarded as a bridge to the customer, research on smart energy services is scarce. To address the gap, we assess smart energy services discussed in research and in the German consumer market and compare the findings from literature with the real market. Our survey provides researchers and practitioners with an overview of smart energy services and can serve as a starting point for service design, which in turn can support the diffusion of energy saving technologies.

Keywords: smart energy services, smart grid services, smart services

1 Introduction

The energy industry is undergoing a transformation process resulting from various technological and economic developments [1]. The transformation is characterized by the expansion of Decentralized Energy Resources (DER) such as Photovoltaic (PV) systems and the emerging digitalization with smart technologies. With several directives (e.g. (2006/32/EC), (2009/72/EC), (2012/27/EU)) the European Union supports this transformation of the energy sector by market liberalization, investments in renewable energy, in smart energy infrastructures and services. As a result, the hierarchical structured power grid transforms towards an intelligent and decentralized grid, the so called smart grid [2].

Against the background of ongoing digitization with the smart grid, new digital energy services emerge which are referred to as smart energy services [3] and promise both new business opportunities and new values for customers [4]. Smart energy services build on smart energy products [4]. Smart meters, for instance, are smart energy products leveraging smart energy services such as energy consumption visualizations, and thus support customers to reduce their energy consumption. Smart energy products and services can not only improve energy efficiency, but can also

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

enable private households to become energy producers with microgeneration units ("prosumer") [5]. Furthermore, smart energy is expected to enhance the options for product differentiation by building on data and consumer usage pattern [6]. Since smart products include smart services by nature [7], research expects a strong shift from product-oriented towards service-oriented offerings [5, 8]. In this regard, a consumer would not need to own the product or technology any more, but would rather use the service [9].

Although smart energy services appear promising and are attributed much significance for the energy domain (e.g. [10]), there is a lack of research on energy services [11], and insights on smart energy services are even more scarce. In terms of smart energy services, research has mainly investigated single services in the past such as smart metering (e.g. [12]) and specific aspects such as willingness to pay (e.g. [13]). Until now, research has provided few overviews on different market available smart energy services for private customers (e.g. [14]). Research has not yet compared conceptualizations of smart energy services found in the literature with smart energy services that exist in a market (particularly in terms of covering a broad range of different smart energy services). Furthermore, there is less research on smart energy services through the lens of Information Systems (IS) and (smart) service science [3]. Against this background, we aim to answer the following research question: *How are existing conceptualizations of smart energy services captured by commercially available smart energy services for private households?* We answer the question by conducting a literature review and analyzing the German market of smart energy services. With this market study approach, we were able to compare the findings from literature with the real market, which is a further step towards existing research. Since smart energy services are eco-efficient by their nature [3], this study also contributes to the field of Energy Informatics [15]. Concerning the practical contribution, the survey provides companies with an up-to-date overview on smart energy services and can serve them as a first step when trying to design new services.

The remainder of this paper is structured as follows: The second section introduces relevant background information on smart energy services. After explaining our method, we present the services derived from the literature review and market analysis. We compare and discuss the findings in section six and seven before drawing a conclusion in the following section.

2 Smart Energy Products and Services

Kranz et al. [3] define smart energy "as the use of ICTs in energy generation, storage, transmission, and consumption, aiming at increasing efficiency, encouraging eco-friendly behavior, and decreasing the emission of GHG [Greenhouse Gas]" (p.8). Moreover, Lund et al. [16] consider smart energy systems as the broader concept in contrast to smart grid, which is an ICT enhanced intelligent grid, and thus, is a part of an overall smart energy system. Moreover, smart energy refers to several kinds of energy not only electricity [13]. A smart energy system can also exist on a household level (e.g. by using a PV-system, storage and intelligent energy manager), and thus does not necessarily need to be connected to the overall power grid [17, 18]. On the household level, smart energy involves smart energy products such as smart meters and intelligent battery storages. Smart energy products consist of physical objects embedding intelligent components, i.e. sensors, controls, software, micro-processors, data storage as well as connectivity, which enable monitoring, control, optimization and autonomous capabilities [4, 19]. Smart services are services made possible by intelligent products, since they use the generated and collected data from smart products, the user and the environment to create new and enhanced customer values such as more convenience and individualized offerings [4, 19]. Following this, smart energy services are an integrated aspect of smart products blurring traditional distinctions between goods and services [7]. Smart products are often regarded as product-service systems, since they combine physical products and digital services as a single solution to the customer [8, 20]. Apart from the use of single smart energy products, smart energy services are enabled by the combination of different integrated smart household appliances, PV-systems, storages, electric vehicles (EVs), etc. and the combination of the data they deliver and their controlling capabilities. Since smart energy services require smart products, which in turn frequently include supporting services, e.g. the financing and installation of DER (such as a PV-system and storage), we would also summarize these services in combination with smart products as smart service offerings [12, 14, 21]. In the following, we analyze smart energy services with a focus on end consumer values and by drawing on the concept of smart products and services (e.g. [4, 19]). Thus, smart energy services can be offered as a pure digital service (e.g. monitoring of energy consumption), but can also be an embedded part of a smart product (e.g. a smart light bulb) and include optional (physical) support services (e.g. installation by a technician).

3 Methodology

To assess a broad range of different smart energy services, we conducted a literature search followed up by a market analysis. For the literature review, we oriented ourselves to established guidelines [22, 23] to ensure a proper procedure and to find a comprehensive sample covering the most important smart energy services. As a first step in identifying relevant articles, we used different keyword combinations of the terms “service”, “services”, “smart energy”, “smart grid”, “smart grids”, “demand side management”, “smart meter”, “smart metering” and “smart home” to search common databases (EBSCOhost, Web of Science, Scopus) between February and April 2018. We restricted our scope to peer-reviewed articles from 2010-2018 and excluded literature not covering the private household sector. As we were primarily interested in literature describing, analyzing and classifying smart energy services from a (smart) service science understanding (e.g. [19, 24]), we did not consider strong technical papers, e.g. describing only system architectures or web-services without considering end consumer services. We also scanned the references (backward search) of proper texts and looked for articles citing these publications (forward search) [22]. After filtering the literature, we identified 28 relevant articles. Consequently, we analyzed the papers by trying to identify services that fulfill our understanding of smart energy services [4, 19]. We grouped related services into clusters by using a combination of deductive and inductive category building in an iterative way [25]. For instance, we grouped services which are often regarded as eco-feedback services in the category monitoring & guidance, whereas other services enable controlling and further autonomous services regarding the energy consumption [4, 9]. Where appropriate, we adopted some already proposed groups in literature, e.g. energy community services and e-mobility services [18, 26].

While the literature review sheds light on the theoretical as-is state on different smart energy services, the aim of the market analysis is to identify a wide range of existing smart energy services for private households in the consumer market. Since the overall research project is embedded in Germany, the search was limited to the German market. Various sources (e.g. rankings of energy companies, blogs, news portals, keyword search in Google) were searched to find a comprehensive amount of German companies offering smart energy services until we noticed a saturation in the findings. After the identification of the companies, we looked specifically for smart energy services on the company's websites. While analyzing the services [25], we tried to apply the service categories from our literature review. However, since the services are often product-oriented and bundled, we differentiated the services as the companies offer them to the customer. This means that one existing smart energy service offering could comprise several single services we identified before according to literature. Furthermore, we only considered services with enough information found on the company site (e.g. information on "smart" features). Since the boarders between different domains become increasingly blurred, we focused on services with a strong connection to energy. For example, we included smart home offerings with a focus on energy and EV charging services but did not include manufacturer of EV. After excluding 29 services not fulfilling our search criteria (e.g. smart home without energy efficiency use cases), our database from the market analysis consisted of an initial sample of 114 smart energy services. Since we were interested in the distinctive elements of different smart energy service offerings and many service providers provide similar offerings, we grouped together similar services.

4 Literature Review of Smart Energy Services

Based on the analysis of the literature, we grouped the smart energy services in superior service categories, which we describe in the following.

Energy Supply & Billing Services (ES&B): Energy supply as a service is the guarantee to supply a specific level of heat, lighting, cooling, which can be facilitated by the collected data (e.g. from smart meter) [11, 27]. In this regard, Pitì et al. [28] state that contracts can be customized based on the measurements. Remote and real-time metering can offer current, e.g. monthly, more accurate and informative billing based on smart metering data [9, 10]. Remote meter reading is also a new service, since no technician needs to get to the household's home to read the meter any more [10].

Monitoring & Guidance (M&G): In this category, we summarized smart energy services that use the data of smart metering and other smart appliances for monitoring the energy consumption and eventually production of private households. The resulting services can offer a wide range of real-time consumption information, feedback and advice. Monitoring and visualization of energy consumption can support consumers to save energy by identifying inefficient behavior and energy guzzlers [9, 13]. When combined with DER a service can also monitor and visualize the own production level [13]. Other services are appliance level energy monitoring, alarm notifications regarding irregular consumption, historical comparison, forecasting, simulation or the comparison of energy consumption with others [9, 13, 29]. Gamified energy consumption information and serious games where participants have to achieve challenges, earn points and can compete with others can also be combined with social media/ social comparison services [5, 30, 31]. Community level feedback provides users with feedback on a community level, e.g. when the neighbor's PV-system is available, or feedback on the contribution users make to the overall system [9]. More concrete feedback and guidance for energy efficiency is provided by setting and monitoring of goals combined with instructions given on how to best achieve certain objectives [21, 32]. Personalized real-time information combined with individual human advice for certain users, devices and user contexts, are particularly interesting for private households [29]. Personal consulting could even include to install the proposed energy efficiency measures [32]. Since the installation of energy efficient measures needs to consider the people and their consumption behavior, we assigned energy efficiency contracting in this category. Regarding this service an energy service company (ESCO) profits from the energy savings [5] in exchange for efficiency measures.

Control & Automation (C&A): Control and automation services enable private households to control smart energy products remotely [9, 21, 33]. Instead of direct control, advanced control mechanisms allow using automatic controls such as using pre-defined rules and settings, e.g. time-setting and configuration of scenarios [33]. By using artificial intelligence some services facilitate self-learning autonomous systems based on user pattern and optimize the energy consumption [34].

Energy Trading (ET): An energy company can undertake smart energy trading services to market the energy produced by consumers [12]. Energy brokering enables energy trading in online marketplaces by automated agents that act for the customers and consider, e.g. user behavior, his calendar, statistical information [6, 35, 36].

Demand Response (DR): Demand response (DR) services aim to induce behavioral change to shift loads to times when renewable energies are available or to hinder peak loads and thus help to stabilize the grid [13]. DR services can comprise incentives that are given to consumers for enabling the utility to shut off household's appliances. Other DR services send signals to the consumers who respond on their own to shift their loads in exchange for financial compensation. Flexible prices (real-time prices, time-of-use pricing etc.) can be further used for load shifting [6, 13, 14].

E-Mobility Services (EM): In the context of e-mobility Niesten and Alkemade [14] name EV batteries to offer "Vehicle to Grid" (V2G) and "Grid to Vehicle" (G2V) services. Further smart energy services are "Vehicle to Home" (V2H)/ "Vehicle as storage" and "Vehicles for DR" [18]. V2G means the use of a fleet of plugged-in EVs as power sources for the ancillary services market and as energy reserves for wholesale markets. Regarding G2V service the consumer charges the EV and pays for the electricity that is charged. V2H uses the storage of the EV to save energy and can be used within the Energy Management System (EMS). "Vehicle for DR" shifts the EV charging process at times of low energy demand or interrupts the charging due to grid instability [18]. Another smart energy service is the provision of EV charging solutions, which can include the sale, installation and management of charging points [32]. In our understanding this would imply smart charging technologies, e.g. a charging station which can respond to price signals and/ or can be integrated to the EMS [18, 28, 37].

Energy Community Services (EC): Smart energy services for community management support not only the pure exchange of information but also the share of energy resources, such as microgeneration units and storages. Private households can register financially or with their capacities such as energy storages and generation plants in a community. The smart energy services organize the generated energy intelligently to balance supply and demand within the community [9, 11]. The energy community services are also referred to as Microgrids (MGs) and Virtual Power Plants (VPPs). In contrast to MGs, VPPs do not need to be locally and physically connected to the smart grid but are virtually aggregated and thus are not bound to a local region [11]. MGs and VPPs can be used for aggregation purposes, i.e. the produced energy of the distributed small generation plants is bundled and is marketed [32].

Smart Home & Smart Metering Set-Up & Support Services (SH&SM): Although not smart in a narrow view, smart energy products must be installed at the customer site first before smart services can be used. In this regard, basic services are the offering of smart home (SH) and smart meter (SH) infrastructures, e.g. via an online marketplace [12, 32]. By SH, we understand smart home systems as part of the home infrastructure, Do-It-Yourself SH systems and corresponding smart electricity appliances from lighting, heating system, washing machine, air-condition, etc. Supporting services can consist of consulting, installing/ integrating, financing and technical support during usage [12, 21, 32]. During system operation the manufacturer can offer remote and predictive maintenance and security services, e.g. by monitoring the status of the appliances remotely, deploying upgrades to the firmware, detecting potential failures and notifying the customer or planning maintenance actions proactively [38].

Decentralized Energy Resources Set-Up & Support Services (DERS): DER can include different microgeneration types such as PV-systems, hybrid heat pumps, micro-combined heat and power (CHP) systems, wind turbines or mini gas turbines. Energy storages are also subsumed under this category, since they are primarily referred to in the context of microgeneration and interesting for prosumers. Especially for DER, supporting services are needed, which can have different contract depths from DER planning, installation/ integration, operation, maintenance and/ or financing. As an alternative for sale of DER, there are options for renting a DER, [11, 26, 32] or infrastructure contracting [26, 32]. Smart services can support asset management in terms of predictive maintenance, intelligent fraud detection and risk management.

Accordingly, the system can detect risks based on data and can take intelligent actions such as switching off assets or sending messages due to irregular events of assets [34].

Integrated Energy Management (IEM): An integrated energy management builds on monitoring and controlling capabilities [4] and intelligently manages and optimizes the consumption, storage and the production autonomously, e.g. in terms of feed-in of excess power and additional energy demand from the grid. The sophisticated services connect numerous production- and consumption-oriented smart energy technologies (e.g. EV, PV-systems, storages, household appliances) and consider diverse context, usage and external data (e.g. tariffs, user preferences, and weather conditions) [28, 34]. We regard an integrated energy management to take place primarily on a household level to optimize the energy flows within the own "smart energy home", but it can also contribute to the overall grid.

5 Market Analysis of Smart Energy Services

Building on our market sample of services, we formed groups according to the offerings made by the service providers, which we explain in the following.

Innovative Energy Supply Tariffs: Innovative energy tariffs are based on smart metering data. For example, Stadtwerke Bielefeld[1] offer a flexible tariff with a smart meter which divides the day into six different price categories.

Smart Home Packages: SH appliances with a focus on energy efficiency are bundled. Basic SH systems consist of a central hub/ gateway and a mobile Application (App), which enable monitoring and controlling features of SH products. Energy efficiency packages (e.g. "Smart Home Paket Energie"[2] by Innogy) bundle different components such as smart lighting, smart heating, smart plugs and different sensors and actuators. The "Viessmann ViCare App"[3], which is included when buying a Viessmann heating systems, enables not only App based automatic control of heating, e.g. scenarios away/ home, but also offers remote maintenance through the service provider.

Digital Add-Ons: In most cases the service provider supplies the physical products with the digital smart service included. However, Innogy sales the App "Storage vario control"[4] for monitoring and controlling its storage separately. Furthermore, on the website of Innogy the consumer can buy add-ons for their Innogy "Smart Home App". For example, the service "Premium-Auswertungen"[5] promises extended functionality for SH through the configuration of scenarios and visualization of consumption data, consumption in € and historic comparisons.

Smart Meter Packages: Another package are smart meters offered together with an App or web portal access for monitoring of energy consumption and further energy

---

1 https://www.stadtwerke-bielefeld.de/privatkunden/tarife/strom/enerbest-strom-smart.html

2 https://www.innogy.com/smartstore/SmarthomeCatalog/SmartHome-Paket-Energie-zid11168110

3 https://www.viessmann.de/de/viessmann-apps/vicare-app.html

4 https://www.innogy.com/smartstore/AppCatalog/Storage-Vario-Control-zid10175961

5 https://www.innogy.com/smartstore/SmarthomeCatalog/Apps/Premium-Auswertungen-zid70011714

efficiency guidance based on smart meter data. As an example, Discovergy[6] offers monitoring and visualizing of real-time, historical and appliance level consumption data. They further provide data-based personalized services in terms of energy reports (e.g. cost control, comparison with others), energy saving tips, tariff consulting and predictive maintenance through analysis of devices and plants to identify defects, consumption analysis with notifications in case of irregular consumption.

Storage Packages: Smart storages are not only provided in combination with PV-systems but also offered as single systems in combination with Apps for monitoring and controlling the storage (e.g. "Voltstorage Smart Storage"[7]). A very innovative service is the "Caterva-Sonne"[8], a smart battery storage. By intelligently networking all storages to a virtually connected intelligent huge storage, the "Caterva-Sonnensystem" provides flexibility to the grid. In exchange for the flexibility, the "Caterva-Sonne" owners earn a financial bonus.

Cloud Storage Services: Another service offering are energy cloud storages, such as the "SolarCloud"[9] by E.ON, which are virtual storages for solar power without the need to own a physical storage. The cloud storages primarily address prosumers with a PV-system and without a physical storage. They can save surplus energy produced by the PV-system in the cloud and get it later when needed.

EV Charging: In terms of e-mobility, the most identified service is the sale of EV charging stations including installation and an App for monitoring the EV charging process. In case of the EV charging station by smartRED[10] they further provide a billing tool, e.g. for billing guests and the integration with a PV-system in order to adapt the charging process to the availability of the produced electricity. Closely related to charging at home is the provision of a public charging infrastructure which is, for example, provided by Innogy and can be accessed with the "eCharge App"[11].

PV-System Bundles with (optional) Storage: Energy companies frequently offer bundles of a PV-system together with a storage and energy management system. It can include the installation, integration and further support services such as maintenance, additional guarantees and insurances. The smart services reach from simple monitoring of energy production and consumption (e.g. RheinEnergie AG "Solarkomfort"[12]) to an intelligent energy management to optimize the level of self-sufficiency (e.g. "E.ON Aura Manager" with PV-system/ storage[13]). Some systems even integrate and control SH appliances and EV charging intelligently for load shifting and further efficiency improvements. Accordingly, a good example for an integrated energy management

---

6 https://discovergy.com/functions

7 https://voltstorage.com/service-und-voltstorage-app/

8 https://www.caterva.de/?produkte#caterva-sonnen

9 https://www.eon-solar.de/eon-solarcloud

10 https://www.smartred.de/e-mobilitaet

11 https://www.innogy.com/web/cms/de/3813766/fuer-zuhause/elektrisch-fahren-und-laden/produkte/produkte-fuer-unterwegs/app-echarge/

12 https://www.rheinenergie.com/de/privatkunden/energieloesungen/photovoltaik_anlage_pachten/photovoltaik_anlage_pachten.html

13 https://www.eon.de/de/pk/solar/aura/manager.html

solution is the "SWM Energiemanager"[14], which is provided in a PV-system and storage full-service bundle. It monitors the energy consumption, enables the control of household appliances with an App, provides recommendations for consumption activities and automatic control, e.g. via time-settings. Further, it considers the energy production, is self-learning and creates a revenue forecast for the PV-system and for the individual consumption. Based on forecast data, it optimizes own consumption and production to increase self-sufficiency, e.g. by scheduling and switching off- and off integrated appliances (e.g. dish washer, heat pumps, EV charging stations). The bundles can also include peer-to-peer (P2P) marketplaces, energy communities, cloud or battery-based flexibility services.

Energy Management Systems: EMSs are likewise offered as single software solutions with a central hub/ gateway and thus come without a PV-system or other smart energy products, e.g. "gridBox Energiemanager". They integrate different smart energy products to monitor, control and optimize (autonomously) the energy flows within the home.

Asset Management: We also identified single specialized smart services for asset management. By automatic remote monitoring "GreenSynergy"[15] provides predictive maintenance and asset performance optimization, e.g. detection of pollution and clouding which minimize the performance. The further deliver status reports as a service for owners of PV-systems.

Energy Communities: Prosumers with and partly without a storage can take part in an energy community in which surplus energy is fed into a "virtual cloud" or directly shared with other community members. Concerning storage provision, the energy community service offers additional benefits such as a free electricity flat rate. For example, the "sonnenFlathome" provides electricity for free, which comes from the community and is refinanced through provision of flexibility for power stabilization. The participation in the community costs a monthly contribution and requires a "sonnen" battery for the virtual network. In case of surplus energy in the "myEnergyCloud"[16] the service provider EWE markets the energy and the revenues are given back to the community. A vendor-independent open approach is the "Energiegruppe" [17] by buzzn GmbH that is open to consumers without a PV-system as well. This energy community is focused on local energy groups, e.g. multi-family houses with a PV-system on the roof. Furthermore, it assists distributed energy groups with the sharing of own produced energy with other energy groups, e.g. with friends or neighborhoods, and in exchange getting produced energy of others. The service provider takes over necessary administrative processes, e.g. billing, regulatory approvals, and change of tenants.

P2P Marketplaces: Energy communities that enable prosumers to buy and sell energy on their own (e.g. defining prices on their own), can be regarded as marketplaces. Digital peer-to-peer energy market platforms directly link decentralized

---

[14] https://www.swm.de/privatkunden/m-strom/photovoltaik/energiemanager.html
[15] https://www.greensynergy.de/anlagenbetreiber/#block1098
[16] https://www.ewe.de/energy-cloud
[17] https://www.buzzn.net/

generation and consumption ("peer-to-peer"). To be financial attractive, P2P marketplaces require highly automated processes, e.g. with blockchain technology to reduce administrative cost. One example here is "OEEX"[18], a blockchain based P2P energy trading platform.

6 Comparison and Analysis of Findings

The literature review and market analysis allowed us to identify smart energy services, which are currently discussed in literature and are already commercially available. To better illustrate and compare the findings, we wrote the number of occurrences of smart energy services in brackets behind the corresponding service types:

The literature frequently concentrates on flexibility and grid stabilization with *Demand Response* (20), but even more relates to energy *Monitoring & Guidance* (22). *Control & Automation* is also a group of services regularly named in the assessed papers (18). Less research is concerned with *Energy Trade* (10), *Energy Community Services* (8) and *Integrated Energy Management* (4).

Although the smart grid is not yet available in Germany, many smart energy services have already entered the German market. In this regard, the market analysis revealed a focus on offering smart products, particularly *SH Packages* (53). Another major part of smart energy services comprises *PV-System Bundles with (optional) Storage* (18), which is not surprising, since the German government has introduced the energy transition away from fossils to renewables and promotes this heavily. We only identified one *Innovative Energy Supply Tariff,* which is a flexible tariff. Furthermore, few offerings are made in terms of pure digital services such as *Digital Add-Ons* (2), *Asset Management* (1), *Cloud Storages* (2), or stand-alone *Energy Communities* (6) and *P2P Marketplaces* (2).

One reason for the product-oriented strategy by companies could be that the domain is still in its infancy and many households first need to have the smart products, such as a smart meter, before being able to buy or consume "pure" (i.e. digital) smart energy services. Moreover, it might not be financially attractive to offer small digital add-ons for private households.

To better contrast the findings from the literature with the market analysis, we assigned the categories from the literature review to the offerings (Table 1). Table 1 shows that most of the identified commercial services relate to *Monitoring & Guidance, Control and Automation, SH & SM Set-Up & Support Services* and *DER Set-Up & Support Services*. Among similar services in one group, different levels of maturity in terms of monitoring, controlling and autonomous actions can be identified, which depends on the integrated components and software algorithms. Above all, services providing control functions such as SH devices normally deliver monitoring functions as well [4].

Our literature review also revealed a fine granular perspective on smart energy services. For instance, we identified several atomic services for energy consumption

[18] http://www.oeex.org/

information (e.g. historic comparisons), which we summarized under the category *Energy Monitoring & Guidance*. In practice, these atomic services are not offered individually, but are also bundled to a comprehensive solution (e.g. smart meter or smart home App offering energy visualizations, historic comparisons, remote control options), which are often provided for free in a smart product bundle. One reason for this can be that literature is not focused on monetization aspects, and thus investigates single smart energy services. Moreover, some services from literature can hardly be monetized through the consumer, since they rather provide process optimizations and improve customer satisfaction, e.g. remote smart meter reading.

Table 1. Service categories mapped to service offerings (numbers in brackets are the frequency of occurrence in the market analysis; abbreviations correspond to the categories from literature)

Market Offering	*Smart Energy Service Category*	*Exemplary Market Offering*
Innovative Energy Supply Tariffs (1)	ES&B, M&G, DR, SH&SM	Stadtwerke Bielefeld Enerbest Strom Smart
SH Package (53)	M&G, C&A, SH&SM	Innogy “Smart Home Paket Energie”, Viessmann ViCare App
Digital Add-Ons (2)	M&G, C&A	Innogy “Premium Auswertungen”
SM Packages (9)	ES&B, M&G, SH&SM	Discovergy Smart Meter,
Storage Packages (8)	M&G, C&A, DERS, EC	Voltstorage Smart, Caterva-Sonne
Cloud Storages (2)	M&G	E.ON SolarCloud
EV Charging (7)	M&G, C&A, EM	smartRED EV Charging station, Innogy eCharge App
PV-System Bundles with (optional) Storage (18)	ES&B, M&G, C&A, DERS, EC, ET, IEM	RheinEnergie AG SolarKomfort, SWM Energiemanager, E.ON Aura Manager Leipziger Sorglos-Servicepaket, ENBW solar+
EMS (5)	M&G, C&A, IEM	gridBox Energiemanager
Energy Communities (6)	M&G, ET, EC, ES&B	EWE myEnergyCloud, sonnenFlathome, buzzn
P2P Marketplaces (2)	ET, IEM	OEEX
Asset Management (1)	DERS	GreenSynergy

When comparing the literature-based conceptualization with the existing services in the market, it becomes clear that some smart energy service offerings introduced in literature are up to now only existing theoretically, e.g. due to regulations, lack of smart grid infrastructures or financial attractiveness. For example, in terms of EV charging there is currently no provider who supports V2G services [18]. Similar demand response services are often not possible, and moreover are not economically feasible due to fixed energy prices [13]. Moreover, we could not find companies offering innovative *Energy Supply & Billing Services* such as tailored contracts or “energy as a service”. This can be due to the domain maturity, since in Germany smart meters are not widespread. Even though no company seems to offer demand response programs in terms of demand shifting, flexibility in forms of resource aggregation (e.g. the storages of Caterva GmbH) has already found its way in the consumer market. Moreover, the flexibility services are not offered separately, but are bundled in a comprehensive product-service offering. Intelligent EMS such as the "SWM Energiemanager" can provide DR on a micro-level by shifting loads to times of higher energy production of the integrated PV-system and storage. Such solutions enable high levels of energy self-sufficiency from the superior power grid and energy utilities. In this regard, "smart home grids" are already realized. Moreover, a commercial service we did not identify in our literature sample is the cloud storage.

7 Discussion

Since many smart energy services are embedded in product-oriented offerings, consumers often face high upfront costs. Apart from classical financing options (e.g. renting, leasing), attempts for a stronger (smart) service orientation can be seen in offering *Digital Add-Ons, Cloud Storages* and allowing consumers to participate in DER in form of communities without owning a PV-system. Digital add-ons can include lower prices for the basic product which then can be updated with additional content for an extra fee. Companies further try to lower adoption barriers by integrating more attractive smart energy services in smart products such as the smart battery storage of Caterva offering consumers extra revenues. Moreover, recent practice projects to study V2G under real conditions are being conducted in Germany, which could also support the diffusion of EVs in the future.[19] Another option to create more attractive offerings are combinations of different smart energy products and services. For instance, an EV can be more valuable if the EV battery storage can be used together with the PV-system in a household (V2H) [18]. Since this is often not yet realized in practice[20], there are promising options such as to use a cloud storage (e.g. E.ON SolarCloud) which can store the own produced energy. The stored energy in the cloud can then be accessed by public charging stations of the service provider. In this regard, (IS) research could study different bundles of smart energy offerings and their acceptance by consumers.

Moreover, to reduce the complexity of smart energy products, physical supporting services are needed, especially to integrate different products, e.g. PV-systems, smart appliances, and EV which is highlighted by the widespread of existing full-service bundles in the market. In this regard, companies should try to develop open systems which can integrate products from several manufacturers. Open systems could also facilitate more stand-alone digital offerings such as digital add-ons or energy communities. Furthermore, complexity is reduced by emerging "plug & play" systems such as PV-systems[21], which are moreover interesting for tenants. To lower financial barriers, literature refers to contracting of smart products and services and usage-based

[19] https://www.electrive.net/2018/10/23/nissan-startet-mit-partnern-v2g-projekt-in-deutschland/

[20] First attempts are shown by the Nissan Leaf with the service "leaf-to-home": https://www.nissan-global.com/EN/TECHNOLOGY/OVERVIEW/vehicle_to_home.html

[21] http://deal.yello.de/mini-solaranlage/

payment modes such as refinancing in terms of energy savings or certificates on the private household level [5, 14] which could be tested and implemented in practice.

Data-based product differentiation and personalized smart energy services are only partly realized. Research could study both innovative ways for monetization and which data-based services (e.g. product recommendations for replacement purchases for old household appliances) could be attractive under which conditions (e.g. privacy, willingness to pay) for end consumers. Due to the forthcoming smart meter roll-out in Germany, more flexible tariffs and demand response programs could emerge for which research could conduct studies with German households to identify promising services. Since we concentrated the market analysis on Germany, there might be different services available in other countries which could be interesting for adopting in the German market as well.

8 Conclusion

Due to the digitalization in the energy industry, smart energy services appear promising for new business opportunities and customer values. Nevertheless, research on smart energy services is scarce. To shed more light on this, we reviewed the literature and the German market and identified not only important characteristics of smart energy services but also the differences in the conceptualizations in literature from the current market offerings. The overview of the as-is state of smart energy services in literature and in the German market landscape might help practitioners to find white spots for new service development and ways how to position their services best. Since smart energy services foster a more sustainable energy production and higher energy efficiency, our work contributes to the IS related Energy Informatics field [15].

One limitation of our paper is the assessment of market available services which relies on website data and is further focused on the German market. Moreover, we did not aim to identify the entirety of literature, but to find a proper sample covering the most important smart energy services. Additionally, we had partly difficulties assessing services to a group during the literature review as several authors do not define, but simply enumerate smart energy services. Accordingly, it also does not always become clear whether the authors refer to “smart” technologies and services. In general, trying to differentiate and describe services is a complex task, since services are partly challenging to isolate [13]. Due to the complexity of isolating single services, not all categories could be arranged without any overlaps.

Notwithstanding the limitations, we are convinced that our survey presents a recent and useful overview on the smart energy service landscape and indicates options for new smart energy service design and fruitful avenues for future research.

References

1. NIST: NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1108r3.pdf.
2. Farhangi, H.: The path of the smart grid. IEEE Power Energy Mag. 8, 18–28 (2010).

3. Kranz, J., Kolbe, L.M., Koo, C., Boudreau, M.C.: Smart energy: where do we stand and where should we go? Electron. Mark. 25, 7–16 (2015).
4. Porter, M.E., Heppelmann, J.E.: How Smart, Connected Products Are Transforming Competition. Harv. Bus. Rev. 1, 3–23 (2014).
5. Hamwi, M., Lizarralde, I., Legardeur, J., Izarbel, T., France, B.: Energy Product Service Systems as core element of energy transition in the household sector : The Greenplay project. In: 22nd International Sustainable Development Research Society Conference (2016).
6. Salah, F., Flath, C.M., Schuller, A., Will, C., Weinhardt, C.: Morphological analysis of energy services: Paving the way to quality differentiation in the power sector. Energy Policy. 106, 614–624 (2017).
7. Fleisch, E., Weinberger, M., Wortmann, F.: Geschäftsmodelle im Internet der Dinge. HMD Prax. der Wirtschaftsinformatik. 51, 812–826 (2014).
8. Valencia, A., Mugge, R., Schoormans, J.P.L., Schifferstein, H.N.J.: The Design of Smart Product-Service Systems (PSSs): An Exploration of Design Characteristics. Int. J. Des. 9, 13–28 (2015).
9. Geelen, D., Reinders, A., Keyson, D.: Empowering the end-user in smart grids: Recommendations for the design of products and services. Energy Policy. 61, 151–161 (2013).
10. Apajalahti, E.L., Lovio, R., Heiskanen, E.: From demand side management (DSM) to energy efficiency services: A Finnish case study. Energy Policy. 81, 76–85 (2015).
11. Hyytinen, K., Toivonen, M.: Future energy services: empowering local communities and citizens. Foresight. 17, 349–364 (2015).
12. Wunderlich, P., Kranz, J., Totzek, D., Veit, D., Picot, A.: The Impact of Endogenous Motivations on Adoption of IT-Enabled Services: The Case of Transformative Services in the Energy Sector. J. Serv. Res. 16, 356–371 (2012).
13. Goldbach, K., Rotaru, A.M., Reichert, S., Stiff, G., Gölz, S.: Which digital energy services improve energy efficiency? A multi-criteria investigation with European experts. Energy Policy. 115, 239–248 (2018).
14. Niesten, E., Alkemade, F.: How is value created and captured in smart grids? A review of the literature and an analysis of pilot projects. Renew. Sustain. Energy Rev. 53, 629–638 (2016).
15. Watson, R.T., Boudreau, M., Chen, A.J.: Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community. MIS Q. 34, 22–38 (2010).
16. Lund, H., Andersen, A.N., Østergaard, P.A., Mathiesen, B.V., Connolly, D.: From electricity smart grids to smart energy systems - A market operation based approach and understanding. Energy. 42, 96–102 (2012).
17. Van Dam, S.S., Bakker, C.A., Van Hal, J.D.M.: Home energy monitors: Impact over the medium-term. Build. Res. Inf. 38, 458–469 (2010).
18. Weiller, C., Neely, A.: Using electric vehicles for energy services: Industry perspectives. Energy. 77, 194–200 (2014).
19. Beverungen, D., Müller, O., Matzner, M., Mendling, J., Vom Brocke, J.: Conceptualizing smart service systems. Electron. Mark. 1–12 (2017).
20. Goedkoop, M.J., Van Halen, C.J.G., Te Riele, H.R.M., Rommens, P.J.M.: Product Service systems , Ecological and Economic Basics. (1999).
21. Richter, L.-L., Pollitt, M.G.: Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market. Energy Econ. 72, 436–450 (2018).

22. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, xiii–xxiii (2002).
23. Vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: ECIS 2009 Proceedings. pp. 2206–2217 (2009).
24. Wünderlich, N. V., Heinonen, K., Ostrom, A.L., Patricio, L., Sousa, R., Voss, C., Lemmink, J.G.A.M.: "Futurizing" smart service: implications for service researchers and managers. J. Serv. Mark. 29, 442–447 (2015).
25. Mayring, P.: Qualitative Inhaltsanalyse. VS Verlag für Sozialwissenschaften. 1, 601–613 (2000).
26. Hamwi, M., Lizarralde, I.: A Review of Business Models towards Service-Oriented Electricity Systems. In: Procedia CIRP. pp. 109–114 (2017).
27. Haas, R., Nakicenovic, N., Ajanovic, A., Faber, T., Kranzl, L., Müller, A., Resch, G.: Towards sustainability of energy systems: A primer on how to apply the concept of energy services to identify necessary trends and policies. Energy Policy. 36, 4012–4021 (2008).
28. Pitì, A., Verticale, G., Rottondi, C., Capone, A., Lo Schiavo, L.: The role of smart meters in enabling real-time energy services for households: The Italian case. Energies. 10, 1–25 (2017).
29. Heiskanen, E., Matschoss, K.: Consumers as innovators in the electricity sector? Consumer perceptions on smart grid services. Int. J. Consum. Stud. 40, 665–674 (2016).
30. Camarinha-Matos, L.M.: Collaborative smart grids – A survey on trends. Renew. Sustain. Energy Rev. 65, 283–294 (2016).
31. Geelen, D., Keyson, D., Boess, S., Brezet, H.: Exploring the use of a game to stimulate energy saving in households. J. Des. Res. 10, 102 (2012).
32. Helms, T.: Asset transformation and the challenges to servitize a utility business model. Energy Policy. 91, 98–112 (2016).
33. Kahma, N., Matschoss, K.: The rejection of innovations? Rethinking technology diffusion and the non-use of smart energy services in Finland. Energy Res. Soc. Sci. 34, 27–36 (2017).
34. Byun, J., Hong, I., Kang, B., Park, S.: A smart energy distribution and management system for renewable energy distribution and context-aware services based on user patterns and load forecasting. Consum. Electron. IEEE Trans. 57, 436–444 (2011).
35. Vytelingum, P., Ramchurn, S.D., Voice, T.D., Rogers, A., Jennings, N.R.: Trading agents for the smart electricity grid. In: 9th International Conference on AAMAS 2010. pp. 897–904 (2010).
36. Karnouskos, S.: Future smart grid prosumer services. In: IEEE PES Innovative Smart Grid Technologies Conference Europe. pp. 1–2 (2011).
37. Immonen, M., Pynnonen, M., Partanen, J., Viljainen, S.: Mapping future services: a case on emerging smart energy metering business. Int. J. Bus. Innov. Res. 4, 491 (2010).
38. Karnouskos, S.: Smart Houses in the Smart Grid and the Search for Value-added Services in the Cloud of Things Era. In: Proceedings of the IEEE International Conference on Industrial Technology. pp. 2016–2021 (2013).

Door-to-Door Mobility Integrators as Keystone Organizations of Smart Ecosystems: Resources and Value Co-Creation – A Literature Review

Thomas Schulz[1,2], Markus Böhm[2], Heiko Gewald[1], and Helmut Krcmar[2]

[1] Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany
{thomas.schulz,heiko.gewald}@hs-neu-ulm.de
[2] Technical University of Munich, Garching, Germany
{markus.boehm,krcmar}@in.tum.de

Abstract. Cities around the world face major mobility-related challenges, such as traffic congestion and air pollution. One primary cause of these challenges is the decision of citizens to use their private car instead of alternative mobility services such as public transport, car-sharing and bike-sharing. Technological progress offers new possibilities to address these challenges by making alternative mobility services easier and more convenient to use. This paper focuses on door-to-door (D2D) mobility integrators, which aim to offer citizens seamless D2D transport by packaging alternative mobility services. To better understand the practical barriers D2D mobility integrators face, this interdisciplinary literature review provides a holistic picture of their operand and operant resources, revealing significant gaps in our understanding of their capability to attract actors to their ecosystem and to manage value co-creation. Based on these gaps, we identify a potential avenue of future research.

Keywords: D2D Mobility Integrators, Literature Review, Operand Resources, Operant Resources, Value Co-Creation.

1 Introduction

Mobility is a basic need of modern society. Currently, most travel is achieved relying on private cars and motorbikes. For example, in Germany motorized private transport constitutes approximately 76% of the modal split of passenger transport [1]. One reason for the prevalence of private motorized transport is the (perceived) weaknesses of alternative mobility services. In the case of public transport, a lack of reliable information about the mobility service and the distance to a station are common perceived weaknesses [2]. Another reason for the prevalence of private motorized transport is that shared services, such as bike-sharing, car-sharing, or ride-sharing have only recently become more comfortable to use due to the increased digital connectivity of the population [3]. As a result, their modal split share has remained relatively low [1].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

However, the impact of such extensive use of motorized private transport is undeniable, especially in cities. Along with an expected rise in the share of people living in cities from 50% in 2015 to 66% by 2050 [4], the number of private cars worldwide is also expected to double by 2035 [5]. The predominant use of the private car causes air and noise pollution, which endangers the health and well-being of the citizens [6], and many cities already face serious traffic congestion and parking problems [7]. An expansion of the road and parking infrastructure is often impossible, prohibitively expensive, or not desirable because natural resources should be sustained. In order to meet these mobility-related challenges, a new mobility paradigm is needed. A promising approach is to combine different alternative mobility services, such as bike-sharing, car-sharing, and public transport, to ease door-to-door (D2D) mobility. This would involve providing up-to-date information about short-term cancellations and delays and adapting trip planning accordingly. If citizens didn't have to gather up-to-date information about multiple alternative mobility service options and undertake a complex comparison/combination and booking process themselves, they may use private cars less [7].

Providing integrated D2D mobility is only possible in an ecosystem of multiple mobility providers, whereby ecosystem is defined as a "system of mostly loosely coupled social and economic (resource-integrating) actors" [8, p. 161]. D2D mobility integrators act as important intermediaries between customers and these mobility providers [9]. Recently, D2D mobility integrators like Moovel (a subsidiary of Daimler AG) and Qixxit (founded by Deutsche Bahn AG) have entered the European mobility market. Their business relies heavily on advanced information technology (IT) and new methods such as business analytics [7, 10]. Analogous to concepts such as 'smart city', 'smart home' [11], or 'smart tourism' [12], the D2D mobility integrators aim to provide a smart service.

Extant research has focused mainly on describing and comparing the quality of different D2D mobility services [7, 10, 13]. Several studies indicate that D2D mobility services provided are often of inferior quality. For example, Albrecht and Ehmke [13] find that D2D mobility integrators struggle to integrate dynamic customer location data and it has been found that only a few mobility providers are willing to join their ecosystem [7, 13]. A valuable theoretical lens through which to investigate the common service provision of multiple actors is service-dominant (S-D) logic [14, 15]. One of the key assumptions of S-D logic is that all actors are engaged in value co-creation – the integration of resources and the exchange of service [15]. To date, however, insufficient attention has been paid to the operand and operant resources actors utilize to provide service in a digital environment [16, 17]. In the case of D2D mobility integrators, there is lack of holistic analysis of their resources and a lack of transparency about resource quality. As a result, it is difficult to appropriately guide the efforts of researchers and practitioners to provide higher quality D2D mobility service. To fill this gap, this study asks the research question:

RQ: What operand and operant resources do D2D mobility integrators utilize to provide D2D mobility services?

To answer this research question, we conduct an interdisciplinary literature review spanning information systems (IS), transportation management, service science and engineering.

The remainder of the paper is structured as follows. After a brief outline of the design of the literature review, we describe our findings. Based on the research gaps identified, we describe an avenue of possible future research. Finally, we discuss the limitations of our study and draw conclusions.

2 Design of the Literature Review

A rigorous and comprehensive literature review, according to Raghuram et al. [18, p. 984] can use "several methodologies such as meta-analysis, descriptive review, and bibliometrics approaches". Each methodology has advantages and disadvantages, giving it a unique "way of seeing [and] a way of not seeing" [19, p. 284]. In this literature review, we follow the guidelines proposed by Webster and Watson [20] in order to include qualitative studies, which are typical for research in emerging fields.

The review process comprises three steps: (1) search leading journals for relevant articles, (2) backward search references in identified articles, and (3) forward search articles referring to the identified articles. In the first step, we searched the Senior Scholars' Basket of Journals of the Association for Information Systems, which contains the leading journals in the IS field. In addition, following the recommendation of Webster and Watson [20], we searched articles presented at the following five important IS conferences – International Conference on Information Systems (ICIS), European Conference on Information Systems (ECIS), Americas Conference on Information Systems (AMCIS), Pacific Asia Conference on Information Systems (PACIS), and Hawaii International Conference on System Sciences (HICSS). Beyond IS, we searched articles contained in the Science Direct and IEEE Xplore databases, which include journal and conference articles published in various fields, such as transportation management, service science, and engineering. Our cut-off date was February 19, 2018. Our keyword search list contained 36 terms in English, including 'intermodal mobility' and 'mobility as a service', as well as the names of well-known D2D mobility integrators (a complete list of keywords can be found in Table 1 of the Appendix).

In addition, we defined the following exclusion and inclusion criteria: (a) we included articles focusing on D2D mobility integrators as part of a digital ecosystem that aims to combine various mobility services. We thus excluded articles dealing with models of transport planning [21, etc.] or physical and organizational transport system integration [e.g., 22]; (b) we excluded articles that focus solely on combining public transport services (e.g., bus, subway, tram) because regional public transport is very often provided by different subsidiaries of the same company [7]. As a result, the parent company faces very different challenges than a D2D mobility integrator cooperating with independent mobility providers; and (c) we excluded theses and books.

Our initial keyword search yielded 4,635 potential articles, which we assessed manually in two rounds. First, we screened the title and, if necessary, the abstract of each article for potential relevance to our research question, selecting 200 articles. Second, we read each remaining article, excluding a further 146 articles. In our backward search, we reviewed the references of the 54 remaining articles, and in our forward search we used the ‘cited by’ function in Google Scholar (https://scholar.google.com). Ten relevant articles could be identified. Furthermore, the anonymous reviewers have proposed one additional article. In all, we identified 65 relevant articles. Figure 1 illustrates our literature review process.

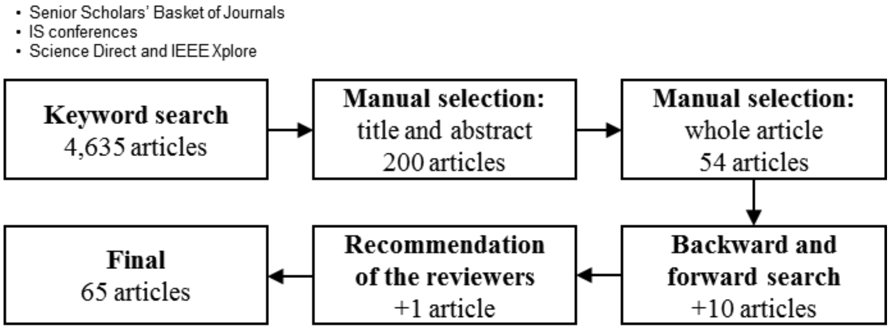


Figure 1. Overview of the literature review process.

We coded the selected articles using an iterative coding approach to ensure internal validity [23]. Our analysis focuses on two main coding categories: operand resources of D2D mobility integrators and operant resources of D2D mobility integrators.

3 Operand and Operant Resources of D2D Mobility Integrators

Operand resources are those “resources on which an operation or act is performed” (e.g., animal life, land, minerals), whereas operant resources “are employed to act on operand resources” (capabilities, competences, knowledge, organizational processes, skills, etc.) [14, p. 2, 24]. In order to provide a D2D mobility service, D2D mobility integrators rely on both types of resources. The classification as operand or operant resource depends on the evaluator’s perspective [25]. Adopting a D2D mobility integrator perspective at the company level, an algorithm for determining the best combination of mobility services is an operand resource. But if the perspective of the algorithm itself is taken, it is an operant resource that adds or does not add a single mobility service to a D2D mobility service. This research adopts the first perspective. The next two sections present the resources of D2D mobility integrators discussed in scientific literature. The identified resources are italicized.

3.1 Operand Resources

Fourteen articles provide a kind of *blueprint* and colloquially describe the basic idea of the new D2D mobility integrator role [26-28]. Frequently, one or more characteristics of the provided D2D mobility service are also discussed in detail. These studies often originated from scientific research projects, as, for instance, in the case of Boero et al. [29], Gogos and Letellier [30], and Motta et al. [31]. Possibly for this reason, the contents of the blueprints differ widely in terms of geographic focus, ranging from a single city [31], to a larger region like Piedmont [32], up to the European Union [29, 30]. Moreover, the authors of the blueprints also adopt differing perspectives in assessing the added value of D2D mobility integrators and their D2D mobility service. Only a few authors [e.g., 30] take a mobility provider perspective, whereas a majority [e.g., 29, 33] take the perspective of citizens and city administrations. Perhaps as a result of these differences, the blueprints take a varying number of mobility services and mobility providers into account [see, e.g., 7, 27, 34]. Nevertheless, they often strive to offer D2D mobility services with similar characteristics, including providing real-time updates and booking and paying for trips through the D2D mobility integrator [35-38].

A further operand resource are *representations* of the *business model* of D2D mobility integrators. Beutel et al. [39] examine how existing business model frameworks are subject of change in the case of D2D mobility integrators and propose a new business model framework for D2D mobility integrators. A similar approach is chosen by Willing et al. [10], who adapt a business model framework to classify multiple D2D mobility integrators. In particular, they highlight the importance of business analytics. Schulz et al. [40] introduce a new model for the roles of D2D mobility integrators based on the roles of an intermediary in the electronic commerce era. Table 2 provides an overview of the operand resources of D2D mobility integrators.

Table 2. Overview of operand resources of D2D mobility integrators.

Operand resource	Authors
Blueprint	[7]; [26]; [27]; [28]; [29]; [30]; [31]; [32]; [33]; [34]; [35]; [36]; [37]; [38]
Representation of the business model	[10]; [39]; [40]
Information system application	[41]; [42]; [43]
National framework architecture	[44]; [45]
Information system architecture	[46]; [47]; [48]; [49]; [50]; [51]; [52]; [53]; [54]; [55]; [56]; [57]; [58]; [59]; [60]; [61]; [62]; [63]
Interface	[64]; [65]
Model and algorithm	[66]; [67]; [68]; [69]; [70]; [71]; [72]; [73]; [74]
Compensation engine	[75]
Recommendation system	[76]; [77]; [78]; [79]

Other articles deal with operand resources which are technical in nature, such as the *information system application*. Based on the trip phases, Digmayer et al. [41] derive the activities of users and the support they require, drawing implications for app design, such as implementing a feedback feature. Stein et al. [42] designed an app, especially for the elderly. Based on their experience in the WISETRIP project, Spitadakis and

Fostieri [43] also make recommendations for improving the design of an app or a web interface.

A related operand resource is ARKTRANS, the *national framework architecture* for mobility information systems in Norway, which encompasses all transport modes and supports D2D mobility [44, 45].

Eighteen articles describe the *information system architecture* of a specific D2D mobility integrator. In line with our previous findings, most of the authors – with exception of Dotoli et al. [46], and Zoghlami et al. [47] – explain the architecture created within the time restraints of their scientific research projects [e.g., 48-51]. It should be noted that several articles focus on the architecture of the Mobility Broker project [52-54], and two articles each focus on the architecture of the IMA [55, 56], Instant Mobility [57, 58], and SMAll [59, 60] project. The authors apply varying architecture styles. For example, Hilgert et al. [61] use the concept of microservices, Motta et al. [62] adopt a service-oriented architecture – event driven architecture (SOA-EDA), whereas Evangelatos et al. [63] draw on the Super Travel API Architecture.

Natvig and Vennesland [64] emphasize the advantage of open *interfaces* for the provision of a D2D mobility service. They also describe the interface definition process and a pilot implementation. Kluth et al. [65] develop an interface that facilitates connecting the rental systems of car-sharing and bike-sharing providers with the information system of D2D mobility integrators. The new interface enables information exchange (e.g., vehicle data, price information) and the execution of bookings via D2D mobility integrators.

A *model* to map the transport system (train stations, bus lines, etc.) and an *algorithm* to identify the best combination of mobility services are further operand resources of D2D mobility integrators. The algorithms can determine the shortest [e.g., 66-69], the fastest [70, 71, etc.], the cheapest [71], the most energy-efficient [72], or the least complex [71] D2D trip. Some algorithms allow the selection from several criteria, while others are restricted to one criterion. In addition, the algorithms vary in terms of whether they work with real-time data [70, 72, etc.] or not [e.g., 71]. Another difference is the scope of mobility services taken into account. For example, Fahnenschreiber et al. [73] and Ma [74] focus on integrating ride-sharing services. Given long running times, the algorithms [e.g., 67, 68, 71] in part prove the basic feasibility of the optimization idea than its practical usability.

Rizzi et al. [75] provide a detailed description of one specific part of the information system, the so-called *compensation engine*, which is responsible for monitoring and re-scheduling the selected combination of mobility services. This technical component enables customers to switch to alternative mobility services in case of a delay or cancellation of a previously chosen mobility service.

While the articles in the last two sections deal with operand resources that are necessary to determine the optimal combination of mobility services based on pre-defined customer criteria, the following articles center on *recommendation systems*, which are deployed to measure and to learn customers' preferences. The framework developed by Samsel et al. [76] helps rate the possible combinations depending on the previously revealed preferences (e.g., tendency to choose the shortest trip), the context (weather, etc.), and the selection made by the crowd. A similar approach focusing on learning customers' preferences is described by Arentze [77], and Zhang and Arentze [78], whose customer criteria include, among others, travel time, monetary costs, environmental impact, and changeover safety. The authors also analyze mobility service preferences (bus, train, etc.). Poxrucker et al. [79] described a simulation tool that can be used to learn from customers' aggregated recommendations and selection. Such an approach is necessary to prevent a bus from overcrowding due to over-recommendation.

3.2 Operant Resources

In order to enable value co-creation, D2D mobility integrators also need different operant resources, such as capabilities, competences and organizational processes, to act on operand resources. Spickermann et al. [80] provide a strategic agenda that highlights the importance of pursuing a *transition strategy*. In particular, they emphasize the need for aspiring D2D mobility integrators to use advanced IT and change their business model. The necessity of a transition strategy is further demonstrated by the work of Sarasini et al. [81] suggesting a research agenda to examine the causes of change and inertia of business models.

Three articles deal with the *implementation process* of a D2D mobility concept. Smith et al. [82] analyze the procurement process of a Swedish public transport organization, which enables a successful bidder to act as a regional D2D mobility integrator. Their results identify seven topics, such as the allocation of responsibilities and technical integration that potential bidders consider important in creating an ecosystem. Khanna and Venters [83, 84] examine the implementation process in the case of the BeMobility project in Berlin. They focus on the D2D mobility integrator's designing an information system for integrating electric car-sharing into public transport infrastructure. Table 3 summarizes the identified operant resources of D2D mobility integrators.

Table 3. Overview of operant resources of D2D mobility integrators.

Operant resource	Authors
Transition strategy	[80]; [81]
Implementation process	[82]; [83]; [84]
Capability to ensure security and privacy	[85]; [86]; [87]; [88]

A further important operant resource of D2D mobility integrators is their *capability* to ensure the *security and privacy* of ecosystem actors. Referring to the SMAll project, Callegati et al. [85] describe the most relevant weaknesses in terms of data reliability, integrity, and authenticity and propose mitigation approaches. For instance, they argue for the implementation of a customer rating system in order to evaluate data sources and enhance data trustworthiness. Further studies deal with insider threats as one of the most prominent security and privacy concerns [86, 87], providing a classification of insider threats. For example, a D2D mobility integrator must be aware of potentially fraudulent data manipulation on behalf of mobility providers. To mitigate insider threats, a networking architecture that based on gossip protocols is introduced. The high relevance of security and privacy is also demonstrated through customer evaluation of the information system of the Mobility Broker project [88].

4 Central Issues for Future Research

This section presents core issues for future research into the operand and operant resources of D2D mobility integrators. We expect that research on these issues will contribute to a deeper understanding of the role of D2D mobility integrators and enable recommendations to put value co-creation into practice.

Our literature review shows that most studies are concerned with the **operand resources** of D2D mobility integrators, such as their information system architecture [e.g., 46, 47], and algorithms for determining the best combination of different mobility services [66, 70, 72, etc.]. We note that D2D mobility integrators should generally be able to access the required operand resources, particularly those of a technical nature. However, many operand resources have only been developed for/used by D2D mobility integrators in the context of scientific research projects. Their real-world practical usefulness remains to be seen. For example, the running times of some of the algorithms [e.g., 67, 68, 71] make them impractical. It is also important to examine how technological progress affects individual technical resources. Currently, the extent to which state-of-the-art IT is taken into account varies greatly. For instance, whereas big data and business analytics are used in the case of the recommendation systems of Poxrucker et al. [79] and Samsel et al. [76], its impact on information system architecture remains unclear [50, 52, etc.].

In contrast, **operant resources** have been largely neglected by scientific research, or, as in the case of pursuing a transition strategy, their importance has only been emphasized in terms of a research agenda [81] or practical implications [80]. In particular, there are no broad insights into the *capability* of D2D mobility integrators *to manage value co-creation* (i.e., the integration of resources and the exchange of service [15]), in the ecosystem. Only three articles [82-84] shed light on the development phase of the ecosystem. The sparsity of research into operant resource needs points to several research gaps, for example what resources are needed in various ecosystem maturation phases, what models of value co-creation are possible in each phase, and how business model transitions can be best managed. Analogously, the phases of the value co-creation process between a D2D mobility integrator and an ecosystem actor (e.g., customer, mobility provider) should also be analyzed.

Since only a few mobility providers have joined an ecosystem of D2D mobility integrators to date [7, 13], future research should provide insights into their *capability to attract actors* – a sub-capability of value co-creation management. In order to explain non-membership, we propose determining the (perceived) value for the actors. As shown in our findings, extant literature predominantly reflects the view of researchers and practitioners adopting the role of D2D mobility integrators [e.g., 30, 37], or working on the development of a single operand resource [56, 65, etc.]. The customer or mobility provider perspective was only seldom considered [42, 43, 88]. However, according to Akaka and Chandler [89, p. 251] (social) roles are important resources for value co-creation "because they guide expectations for the exchange of service". To date, little scholarship is available into how actors evaluate the role of D2D mobility integrators and its related set of practices. One exception is Beutel et al. [88], who find that providing price bundles is often evaluated negatively by customers due to data security and privacy concerns. Overall, we believe that identifying the divergent expectations of actors about the role of D2D mobility integrators can help to better understand what motivates their decision to join or not join a D2D mobility ecosystem.

The capability of D2D mobility integrators to attract actors to their ecosystem depends not only on whether they fulfill the expectations of potential customers and mobility providers. As seen through the lens of S-D logic [14, 15], which assumes that all actors are engaged in value co-creation, every actor is already embedded in multiple ecosystems coordinated by a set of rules, norms and beliefs, known as institutions, and higher-order collections of these institutions, known as institutional arrangements [15]. For instance, German mobility providers that provide bus and tram transport are often integrated into the ecosystem of a municipal utility which is acting as parent company [7]. Hence, the decision of potential customers and mobility providers on membership in the D2D mobility integrator ecosystem is influenced by the actors of their existing ecosystems.

This literature review indicates that no research has been done on this topic. One avenue of possible future research is how the existing institutional arrangements of potential customers and mobility providers influence their decision whether or not to join the ecosystem of D2D mobility integrators. This might be viewed through the theoretical lens of legitimacy [90]. A potential study could consider the interests and authority of all actors of established ecosystems, such as parent companies, city administrations and industry associations in terms of preventing or mandating collaboration with D2D mobility integrators. Relying on power-based theory [91], this approach would help D2D mobility integrators better understand the relevant actors and their interests, and adapt their role and the institutional arrangements established in the ecosystem to increase value co-creation.

In summary, we argue for a two-step analysis to enhance our knowledge about the capability of D2D mobility integrators to attract actors to their ecosystem – the evaluation of the fulfillment of the role expectations by D2D mobility integrators and the actors' embeddedness in existing ecosystems. We assume that this approach can also be used to investigate the capability of D2D mobility integrators to manage value co-creation in the later phases of the ecosystem and better understand why mobility providers, for instance, provide real-time timetable data but are not able or willing to participate in a common ticketing.

5 Conclusion

In this paper, we conduct a literature review on the operand and operant resources required by D2D mobility integrators. Based on our findings, we propose several avenues of future research, including analyzing the operand resources used by D2D mobility integrators outside the project environment, D2D mobility integrators' capability to manage value co-creation in the different phases of the ecosystem, and the sub-capability of D2D mobility integrators to attract actors to their ecosystem in the first phase.

By reviewing scientific literature and deriving avenues of future research, our study contributes to IS and S-D logic literature in several ways. First, it provides a holistic view on research on the operand and operant resources of D2D mobility integrators. Our study integrates previously unrelated studies focusing on a single resource. This overview of the current state of research and, to a degree, practice, enables us to identify research gaps and practical challenges, which serve as the basis for developing avenues of future research. Practitioners can benefit from our work by understanding the resources necessary to put value co-creation into practice.

Our study is subject to some limitations. First, despite our backward and forward searches, our selection of outlets and keywords may have excluded some relevant scholarship. Second, the level of granularity and aggregation we choose in terms of analyzing resources may have excluded some insights into additional resources.

Appendix

Table 1. Keywords.

Allryder	GoEuro	Mobility Map	Multimodal mobility
Citymapper	Integrated mobility	Mobility marketplace	Multimodal travel
Connected mobility	Intermodal mobility	Mobility markets	Networked mobility
D2D mobility	Intermodal Mobility Assistance for Megacities (IMA)	Mobility network	Qixxit
Door 2 Door mobility	Intermodal travel	Mobility on demand	Rome2rio
Door2door	MeMobility	Mobility platform	RouteRANK
Door-to-door mobility	Mobility as a service	Modular mobility	Smart mobility
FromAtoB	Mobility Broker Project	Moovel	Transit App
Future mobility	Mobility ecosystem	Moovit	Waymate

6 Acknowledgements

We thank the Bavarian State Ministry of Science and the Arts for funding this research. Thomas Schulz was also supported by the BayWISS Consortium Digitization. The authors retain full responsibility for the content of this publication.

References

1. Bundesministerium für Verkehr und digitale Infrastruktur: Verkehr in Zahlen 2016/2017 (2016).
2. Beirão, G., Cabral, J.A.S.: Understanding Attitudes Towards Public Transport and Private Car: A Qualitative Study. Transport Policy 14, 478-489 (2007).
3. Puschmann, T., Alt, R.: Sharing Economy. Business & Information Systems Engineering 58, 93-99 (2016).
4. United Nations Department of Economic and Social Affairs: World Urbanization Prospects (2015).
5. International Energy Agency: World Energy Outlook 2012 (2012).
6. Barth, M., Boriboonsomsin, K.: Real-World Carbon Dioxide Impacts of Traffic Congestion. Transportation Research Record: Journal of the Transportation Research Board, 163-171 (2008).
7. Willing, C., Brandt, T., Neumann, D.: Intermodal Mobility. Business & Information Systems Engineering 59, 173-179 (2017).
8. Lusch, R.F., Nambisan, S.: Service Innovation: A Service-Dominant Logic Perspective. MIS Quarterly 39, 155-175 (2015).
9. Agarwal, N., Soh, C., Yeow, A.: Value Co-Creation in Service Ecosystems: A Member Perspective. In: International Conference on Information Systems (2016).
10. Willing, C., Brandt, T., Neumann, D.: Electronic Mobility Market Platforms – A Review of the Current State and Applications of Business Analytics. Electronic Markets 27, 267-282 (2017).
11. Rocznik, D., Goffart, K., Wiesche, M., Krcmar, H.: Towards Identifying User-Centered Requirements for Smart In-House Mobility Services. KI-Künstliche Intelligenz 31, 249-256 (2017).
12. Gretzel, U., Sigala, M., Xiang, Z., Koo, C.: Smart Tourism: Foundations and Developments. Electronic Markets 25, 179-188 (2015).
13. Albrecht, L., Ehmke, J.F.: Innovative Services in der Mobilitätsbranche: Eine Marktanalyse multimodaler Mobilitätsmanager. In: Multikonferenz Wirtschaftsinformatik (2016).
14. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. Journal of Marketing 68, 1-17 (2004).
15. Vargo, S.L., Lusch, R.F.: Service-Dominant Logic 2025. International Journal of Research in Marketing 34, 46-67 (2017).
16. Lenka, S., Parida, V., Wincent, J.: Digitalization Capabilities as Enablers of Value Co-Creation in Servitizing Firms. Psychology & Marketing 34, 92-100 (2017).
17. Brust, L., Breidbach, C.F., Antons, D., Salge, T.O.: Service-Dominant Logic and Information Systems Research: A Review and Analysis Using Topic Modeling. In: International Conference on Information Systems (2017).
18. Raghuram, S., Tuertscher, P., Garud, R.: Research Note - Mapping the Field of Virtual Work: A Cocitation Analysis. Information Systems Research 21, 983-999 (2010).
19. Poggi, G.: A Main Theme of Contemporary Sociological Analysis: Its Achievements and Limitations. British Journal of Sociology 16, 283-294 (1965).
20. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, xiii-xxiii (2002).
21. Hoogendoorn-Lanser, S., van Nes, R., Hoogendoorn, S.P., Bovy, P.: Home-Activity Approach to Multimodal Travel Choice Modeling. Transportation Research Record: Journal of the Transportation Research Board, 180-187 (2006).

22. Li, L., Loo, B.P.Y.: Towards People-centered Integrated Transport: A Case Study of Shanghai Hongqiao Comprehensive Transport Hub. Cities 58, 50-58 (2016).
23. Strauss, A.L., Corbin, J.M.: Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. SAGE Publications, Thousand Oaks (1998).
24. Madhavaram, S.R., Hunt, S.D.: The Service-Dominant Logic and a Hierarchy of Operant Resources: Developing Masterful Operant Resources and Implications for Marketing Strategy. Journal of the Academy of Marketing Science 36, 67-82 (2008).
25. Spohrer, J., Vargo, S.L., Caswell, N., Maglio, P.P.: The Service System is the Basic Abstraction of Service Science. In: Hawaii International Conference on System Sciences (2008).
26. Giesecke, R., Surakka, T., Hakonen, M.: Conceptualising Mobility as a Service: A User Centric View on Key Issues of Mobility Services. In: Conference on Ecological Vehicles and Renewable Energies (2016).
27. Kamargianni, M., Li, W., Matyas, M., Schäfer, A.: A Critical Review of New Mobility Services for Urban Transport. Transportation Research Procedia 14, 3294-3303 (2016).
28. Kamargianni, M., Matyas, M.: The Business Ecosystem of Mobility-as-a-Service. In: Transportation Research Board (2017).
29. Boero, M., Garré, M., Fernandez, J., Persi, S., Quesada, D., Jakob, M.: MyWay Personal Mobility: From Journey Planners to Mobility Resource Management. Transportation Research Procedia 14, 1154-1163 (2016).
30. Gogos, S., Letellier, X.: IT2Rail: Information Technologies for Shift to Rail. Transportation Research Procedia 14, 3218-3227 (2016).
31. Motta, G., Sacco, D., Belloni, A., You, L.: A System for Green Personal Integrated Mobility: A Research in Progress. In: International Conference on Service Operations and Logistics, and Informatics (2013).
32. Arneodo, F., Castelli, R., Botta, D.: Towards a "Smart Region" Paradigm: Beyond Smart Cities Borders: Piedmont Region Experience. In: International Conference of Electrical and Electronic Technologies for Automotive (2017).
33. Motta, G., Ferrara, A., Sacco, D., You, L., Cugola, G.: Integrated Mobility: A Research in Progress. Journal of Software Engineering and Applications 6, 97-101 (2013).
34. Jittrapirom, P., Caiati, V., Feneri, A.-M., Ebrahimigharehbaghi, S., Alonso-González, M.J., Narayan, J.: Mobility as a Service: A Critical Review of Definitions, Assessments of Schemes, and Key Challenges. Urban Planning 2, 13-25 (2017).
35. Deniaud, I., Quiguer, S., Breuil, D., Le Maguet, P., Lecourt, J., Pourcel, C., Ruault, J.-R., Somat, A.: Interoperability Dimensions for Multimodal Mobility Management. In: IFAC Symposium on Information Control Problems in Manufacturing (2012).
36. Melis, A., Mirri, S., Prandi, C., Prandini, M., Salomoni, P., Callegati, F.: Integrating Personalized and Accessible Itineraries in MaaS Ecosystems Through Microservices. Mobile Networks and Applications 23, 167-176 (2017).
37. Beutel, M.C., Krempels, K.-H.: Encompassing Payment for Heterogeneous Travelling: Design Implications for a Virtual Currency based Payment Mechanism for Intermodal Public Transport. In: International Conference on Smart Grids and Green IT Systems (2014).
38. Ambrosino, G., Nelson, J.D., Boero, M., Pettinelli, I.: Enabling Intermodal Urban Transport through Complementary Services: From Flexible Mobility Services to the Shared Use Mobility Agency: Workshop 4. Developing Inter-Modal Transport Systems. Research in Transportation Economics 59, 179-184 (2016).
39. Beutel, M.C., Samsel, C., Mensing, M., Krempels, K.-H.: Business Model Framework to Provide Heterogeneous Mobility Services on Virtual Markets. In: International Conference on e-Business (2014).

40. Schulz, T., Rockmann, R., Weeger, A.: Service Composition in Networks – Towards a Typology of Intermediaries. In: Americas Conference on Information Systems (2016).
41. Digmayer, C., Vogelsang, S., Jakobs, E.-M.: Designing Mobility Apps to Support Intermodal Travel Chains. In: International Conference on the Design of Communication (2015).
42. Stein, M., Meurer, J., Boden, A., Wulf, V.: Mobility in Later Life - Appropriation of an Integrated Transportation Platform. In: Conference on Human Factors in Computing Systems (2017).
43. Spitadakis, V., Fostieri, M.: WISETRIP - International Multimodal Journey Planning and Delivery of Personalized Trip Information. Procedia - Social and Behavioral Sciences 48, 1294-1303 (2012).
44. Natvig, M.K., Westerheim, H.: National Multimodal Travel Information – A Strategy based on Stakeholder Involvement and Intelligent Transportation System Architecture. IET Intelligent Transport Systems 1, 102-109 (2007).
45. Natvig, M.K., Westerheim, H.: Refinement and Evaluation of the ARKTRANS Framework Through Use in Travel Information Services. In: World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting (2008).
46. Dotoli, M., Zgaya, H., Russo, C., Hammadi, S.: A Multi-Agent Advanced Traveler Information System for Optimal Trip Planning in a Co-Modal Framework. IEEE Transactions on Intelligent Transportation Systems 18, 2397-2412 (2017).
47. Zoghlami, N., Jeribi, K., Merlo, C., Zgaya, H., Hammadi, S.: A Multi-Agent Architecture for a Co-Modal Transport System. Journal of Modern Engineering Research 4, 60-78 (2014).
48. Fiorentino, A., De Gioia, C., Gaido, M., Conti, G., Magliocchetti, D., De Amicis, R., Kipp, W.: Mobile Integration Platform Concept: The Naples Pilot Test Site. Procedia - Social and Behavioral Sciences 48, 1855-1864 (2012).
49. Eryilmaz, E., Kagerbauer, M., Schuster, T., Wolf, O.: Collaborative Management of Intermodal Mobility. In: Working Conference on Virtual Enterprises (2014).
50. Solar, A., Marqués, A.: ENHANCED WISETRIP: Wide Scale Multimodal and Intelligent Journey Planning. Procedia-Social and Behavioral Sciences 48, 2940-2949 (2012).
51. Zografos, K., Madas, M.: Optimizing Intermodal Trip Planning Decisions in Interurban Networks. Transportation Research Record: Journal of the Transportation Research Board 61-69 (2003).
52. Beutel, M.C., Gökay, S., Kluth, W., Krempels, K.-H., Samsel, C., Terwelp, C.: Product Oriented Integration of Heterogeneous Mobility Services. In: International Conference on Intelligent Transportation Systems (2014).
53. Beutel, M.C., Gökay, S., Kluth, W., Krempels, K.-H., Ohler, F., Samsel, C., Terwelp, C., Wiederhold, M.: Information Integration for Advanced Travel Information Systems. Journal of Traffic and Transportation Engineering 4, 177-185 (2016).
54. Beutel, M.C., Gökay, S., Kluth, W., Krempels, K.-H., Samsel, C., Terwelp, C., Wiederhold, M.: Heterogeneous Travel Information Exchange. In: International Conference on Mobility in IoT (2016).
55. Masuch, N., Lützenberger, M., Keiser, J.: An Open Extensible Platform for Intermodal Mobility Assistance. Procedia Computer Science 19, 396-403 (2013).
56. Keiser, J., Masuch, N., Lützenberger, M., Grunewald, D., Kern, M., Trollmann, F., Acar, E., Salma, Ç.A., Dang, X.-T., Kuster, C.: IMA - An Adaptable and Dynamic Service Platform for Intermodal Mobility Assistance. In: International Conference on Intelligent Transportation Systems (2014).

57. Zargayouna, M., Zeddini, B., Scemama, G., Othman, A.: Simulating the Impact of Future Internet on Multimodal Mobility. In: International Conference on Computer Systems and Applications (2014).
58. Zargayouna, M., Scemama, G., Zeddini, B., Kompfner, P., Gatellier, P., Constant, P., Beckman, D.: Future Internet for a Personal Travel Companion Service. In: ITS World Congress (2012).
59. Callegati, F., Delnevo, G., Melis, A., Mirri, S., Prandini, M., Salomoni, P.: I Want to Ride My Bicycle: A Microservice-Based Use Case for a MaaS Architecture. In: IEEE Symposium on Computers and Communications (2017).
60. Callegati, F., Gabbrielli, M., Giallorenzo, S., Melis, A., Prandini, M.: Smart Mobility for All: A Global Federated Market for Mobility-as-a-Service Operators. In: International Conference on Intelligent Transportation (2017).
61. Hilgert, T., Kagerbauer, M., Schuster, T., Becker, C.: Optimization of Individual Travel Behavior through Customized Mobility Services and their Effects on Travel Demand and Transportation Systems. Transportation Research Procedia 19, 58-69 (2016).
62. Motta, G., Sacco, D., Ma, T., You, L., Liu, K.: Personal Mobility Service System in Urban Areas: The IRMA Project. In: IEEE Symposium on Service-Oriented System Engineering (2015).
63. Evangelatos, S., Kalampoukis, Z., Fergadioti, I., Christofi, S., Karakostas, B., Zorgios, Y.: Service Availability Analysis of a Multimodal Travel Planner Using Stochastic Automata. In: IEEE Symposium on Computers and Communications (2017).
64. Natvig, M.K., Vennesland, A.: Flexible Organisation of Multimodal Travel Information Services. IET Intelligent Transport Systems 4, 401-412 (2010).
65. Kluth, W., Beutel, M.C., Gökay, S., Krempels, K.-H., Samsel, C., Terwelp, C.: IXSI - Interface for X-Sharing Information. In: International Conference on Web Information Systems and Technologies (2015).
66. Delling, D., Pajor, T., Wagner, D.: Accelerating Multi-Modal Route Planning by Access-Nodes. In: European Symposium on Algorithms (2009).
67. Zhang, J., Liao, F., Arentze, T., Timmermans, H.: A Multimodal Transport Network Model for Advanced Traveler Information Systems. Procedia - Social and Behavioral Sciences 20, 313-322 (2011).
68. Zhang, J., Liao, F., Arentze, T., Timmermans, H.: A Multimodal Transport Network Model for Advanced Traveler Information Systems. Procedia Computer Science 5, 912-919 (2011).
69. Hrncir, J., Jakob, M.: Generalised Time-Dependent Graphs for Fully Multimodal Journey Planning. In: International IEEE Conference on Intelligent Transportation Systems (2013).
70. Rouhieh, B., Alecsandru, C.: Optimizing Route Choice in Multimodal Transportation Networks. Canadian Journal of Civil Engineering 41, 800-810 (2014).
71. Yu, H., Lu, F.: A Multi-Modal Route Planning Approach With an Improved Genetic Algorithm. In: International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science (2010).
72. Prandtstetter, M., Straub, M., Puchinger, J.: On the Way to a Multi-Modal Energy-Efficient Route. In: Conference of the IEEE Industrial Electronics Society (2013).
73. Fahnenschreiber, S., Gündling, F., Keyhani, M.H., Schnee, M.: A Multi-Modal Routing Approach Combining Dynamic Ride-Sharing and Public Transport. Transportation Research Procedia 13, 176-183 (2016).
74. Ma, T.-Y.: On-Demand Dynamic Bi-/Multi-Modal Ride-Sharing Using Optimal Passenger-Vehicle Assignments. In: IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe (2017).

75. Rizzi, F., Motta, G., Sacco, D.: A System for Green Personal Integrated Mobility: Compensation Engine. International Journal of Computer Applications 105, 5-11 (2014).
76. Samsel, C., Garbereder, G., Krempels, K.-H.: Personalized, Context-aware Intermodal Travel Information. In: International Conference on Web Information Systems and Technologies (2016).
77. Arentze, T.A.: Adaptive Personalized Travel Information Systems: A Bayesian Method to Learn Users' Personal Preferences in Multimodal Transport Networks. IEEE Transactions on Intelligent Transportation Systems 14, 1957-1966 (2013).
78. Zhang, J., Arentze, T.A.: Personalized Multi-Modal Route Planning: A Preference-Measurement and Learning-Based Approach. In: International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (2014).
79. Poxrucker, A., Bahle, G., Lukowicz, P.: Simulating Adaptive, Personalized, Multi-Modal Mobility in Smart Cities. In: EAI International Summit Smart City 360° (2016).
80. Spickermann, A., Grienitz, V., von der Gracht, H.A.: Heading Towards a Multimodal City of the Future?: Multi-stakeholder Scenarios for Urban Mobility. Technological Forecasting and Social Change 89, 201-221 (2014).
81. Sarasini, S., Linder, M., Langeland, O., Julsrud, T.-E.: Integrating a Business Model Perspective into Sustainability Transitions: A Research Agenda based on Servitised Mobility. In: International Sustainability Transitions Conference (2016).
82. Smith, G., Sochor, J., Karlsson, M.: Procuring Mobility as a Service: Exploring Dialogues with Potential Bidders in West Sweden. In: ITS World Congress (2017).
83. Khanna, A., Venters, W.: The Role of Intermediaries in Designing Information Infrastructures in Strategic Niches: The Case of a Sustainable Mobility Infrastructure Experiment in Berlin. In: European Conference on Information Systems (2013).
84. Khanna, A., Venters, W.: Exploring the Rhythms of Information Infrastructure Coordination for Smart Cities: The Case of Building a Mobility Infrastructure in Berlin. In: European Conference on Information Systems (2014).
85. Callegati, F., Giallorenzo, S., Melis, A., Prandini, M.: Data Security Issues in MaaS-Enabling Platforms. In: International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (2016).
86. Callegati, F., Giallorenzo, S., Melis, A., Prandini, M.: Insider Threats in Emerging Mobility-as-a-Service Scenarios. In: Hawaii International Conference on System Sciences (2017).
87. Callegati, F., Giallorenzo, S., Melis, A., Prandini, M.: Cloud-of-Things meets Mobility-as-a-Service: An Insider Threat Perspective. Computers & Security 74, 277-295 (2018).
88. Beutel, M.C., Zaunbrecher, B.S., Himmel, S., Krempels, K.-H., Ziefle, M.: Evaluation of an Integrated Intermodal Travel Service. In: International Conference on Smart Cities and Green ICT Systems (2016).
89. Akaka, M.A., Chandler, J.D.: Roles as Resources: A Social Roles Perspective of Change in Value Networks. Marketing Theory 11, 243-260 (2011).
90. Dacin, M.T., Oliver, C., Roy, J.P.: The Legitimacy of Strategic Alliances: An Institutional Perspective. Strategic Management Journal 28, 169-187 (2007).
91. Rodríguez, C., Langley, A., Béland, F., Denis, J.-L.: Governance, Power, and Mandated Collaboration in an Interorganizational Network. Administration & Society 39, 150-193 (2007).

Ein Entscheidungsunterstützungssystem zur ökonomischen Bewertung von Mieterstrom auf Basis der Clusteranalyse

Jannick Töppel[1]

[1] Universität Augsburg, Kernkompetenzzentrum Finanz- & Informationsmanagement, Augsburg, Deutschland
jannick.toeppel@fim-rc.de

Abstract. Für den Erfolg der Energiewende spielt die dezentrale Stromerzeugung eine entscheidende Rolle. Aus diesem Grund wurde das Geschäftsmodell Mieterstrom entwickelt, welches sich über die Erzeugung und Lieferung von Strom in direktem räumlichen Zusammenhang definiert. Dabei soll durch den direkten Verkauf von Strom an Mieter ein höherer Gewinn erzielt werden, im Vergleich zur klassischen Stromeinspeisung ins Netz. Zur Förderung von Mieterstrom in Deutschland wurde 2017 von politischer Seite ein umfassendes Förderprogramm beschlossen und somit die Rahmenbedingungen deutlich verbessert. Basierend auf Smart Meter Daten wird in diesem Beitrag deshalb ein Entscheidungsunterstützungsmodell zur Investitionsbewertung von Mieterstrommodellen entwickelt und evaluiert. Dafür wird in einem ersten Schritt eine Clusteranalyse durchgeführt, um anschließend auf Basis der durchschnittlichen Stromverbrauchsprofile der Cluster sowie bedingter Wahrscheinlichkeiten die Rentabilität eines Mieterstrommodells vorherzusagen. Die in einer Fallstudie evaluierten Investitionen weisen eine Amortisationszeit von 8 bis 14 Jahren sowie eine CO_2-Einsparung von über 60% auf.

Keywords: Mieterstrom, Investitionsbewertung, Clusteranalyse, Smart Meter Daten.

1 Einleitung

Vor dem Hintergrund des voranschreitenden Klimawandels sowie knapper werdender Ressourcen, wurden in den vergangenen Jahren weltweit erste Anstrengungen unternommen, um diesen negativen Entwicklungen entgegenzuwirken [1]. In Deutschland spielt zur Erreichung der internationalen Klimaziele die sogenannte Energiewende eine zentrale Rolle, wohinter sich der kurzfristige Ausstieg aus der Atomenergie und der mittelfristige Ausstieg aus der Kohleenergie verbirgt [2]. Eine wichtige Voraussetzung für das Erreichen dieser Ziele ist der Aufbau einer dezentralen Energieversorgung, für welche der Deutsche Bundesrat am 7. Juli 2017 mit dem Gesetz zur Förderung von Mieterstrom einen wichtigen Grundstein legte [3]. Mieterstrom stammt aus einer Photovoltaikanlage (PV-Anlage) oder einem Blockheizkraftwerk

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

(BHKW) eines Wohngebäudes und wird direkt den ansässigen Mietern zur Verfügung gestellt [4]. Überschüssiger Strom kann gegen die übliche EEG-Einspeisevergütung dem Stromnetz zugeführt werden. Diverse steuerliche Erleichterungen und Zuschüsse machen es für Vermieter nun deutlich attraktiver, den produzierten Strom möglichst vollständig an ihre Mieter zu verkaufen. Um den Mietern die Teilnahme am Mieterstrommodell interessant zu machen, wird in der Regel der Marktpreis für Strom leicht unterboten. Für den Vermieter wird dadurch der Eigenverbrauch, also der direkt im Gebäude verbrauchte Anteil des erzeugten Stroms, zu einer entscheidenden Performancekennzahl. Um die Stromverbräuche und Stromeinspeisungen präzise messen zu können, spielen Smart Meter eine zentrale Rolle in Mieterstrommodellen. Die erfassten, hochauflösenden Messergebnisse werden insbesondere für die komplexe Abrechnung benötigt, welche in der Regel über eine unterstützende IT Infrastruktur abgewickelt wird [5].

Die durch die Gesetzesänderung geschaffene Ausgangslage hat bereits erste Wohnbaugesellschaften zu Investitionen in das Mieterstrommodell bewegt [6]. Deren Rentabilität hängt, neben der Entwicklung des Strompreises, maßgeblich vom Eigenverbrauch im Tagesverlauf ab [7]. Ein Beispiel: Wenn die meisten Mieter tagsüber während der sonnenreichen Stunden nicht zuhause sind, muss der Investor den Großteil des produzierten Photovoltaikstroms ins Netz einspeisen und profitiert daher nicht von den höheren Preisen, welche er von den Mietern verlangen könnte. Ex ante sind, in Abwesenheit von Smart Metern, die Verbrauchsprofile in der Regel nicht bekannt oder nur in aggregierter Form (Jahresverbräuche) verfügbar und können so nicht bzw. kaum zur Investitionsbewertung herangezogen werden.

Auf Basis von Smart Meter Stromdaten wurden in den vergangenen Jahren eine Vielzahl von Data Analytics Lösungen entwickelt. So wurden Haushalte in Abhängigkeit von Verbrauchsprofilen in verschiedene Cluster eingeteilt [8] und anhand des Verbrauchsprofils auf Eigenschaften, wie z.B. das Gebäudealter, die Art der Warmwasserversorgung [9] sowie den Beschäftigungsstatus oder die Wohnform [10], geschlossen. Dabei liegt der Fokus der meisten Arbeiten jedoch nicht auf der Entscheidungsunterstützung durch Informationssysteme (IS) bei der Investitionsbewertung, sondern vielmehr auf der Unterstützung von politischen Entscheidungen oder der Entwicklung neuer Dienstleistungen für Endkunden, Netzbetreiber und Stromanbieter.

Aus den vorangegangen Überlegungen resultiert die folgende Forschungsfrage: Wie kann ein Entscheidungsunterstützungssystem basierenden auf Smart Meter Daten das Verbrauchsprofil einer Immobilie vorhersagen, um Mieterstrommodelle ökonomisch zu bewerten?

Die datenbasierte Vorhersage von Verbrauchsprofilen für Mieterstrommodelle ist ein Beitrag zur Energie Informatik, welche als Teilgebiet der IS-Forschung Informationssysteme zur Steigerung der Energieeffizienz und der Nachhaltigkeit entwickelt [11]. Um die formulierte Forschungsfrage zu beantworten, wurde der Design Science Ansatz gewählt [12]. Dieser Ansatz zeichnet sich durch die Entwicklung und anschließende Evaluation eines Artefakts aus. Hevner et al. [12] heben hervor, dass ein entwickeltes Artefakt (Modell) eine praktische Implikation zeigen sollte. Diese Anforderung wird in Kapitel 2 durch die Einbettung des Artefakts in ein Geschäftsmodell gewährleistet, welches auf Basis einer kurzen Einführung zu Mieterstrom und Datenanalysen von Smart Meter Daten skizziert wird. Anschließend wird in Kapitel 3 ein datenbasiertes Modell zur Vorhersage von Eigenverbräuchen entwickelt, welches zur ökonomischen Bewertung von Mieterstrommodellen verwendet werden kann. In Kapitel 4 wird das entwickelte Modell anhand realer Daten evaluiert und erprobt. Abschließend werden in Kapitel 5 die Ergebnisse und Limitationen zusammengefasst sowie auf weitere Forschungsfragen eingegangen.

2 Mieterstrom und Datenanalyse

Mit sinkenden Kosten für die lokale Stromerzeugung [13] und steigenden Endkundenpreisen für Strom, wird die dezentrale Stromerzeugung und der direkte Konsum vor Ort immer attraktiver. Der wirtschaftliche Rahmen, in welchem ein Investor (Vermieter) von diesen Entwicklungen profitieren kann, ist das oben eingeführte Mieterstrommodell. Behr und Großklos [4] definieren Mieterstrom anhand vier wesentlicher Elemente: (1) Die dezentrale Erzeugung von Strom, welche entweder regenerativ oder mit hoher Effizienz in Kraft-Wärme-Kopplung erfolgt, (2) der Standort der Energieerzeugungsanlage befindet sich in einem Gebäude oder auf dem Wohngelände, (3) zwischen der Energieerzeugungsanlage und den Mietern besteht ein räumlicher Zusammenhang sowie (4) für die Belieferung der Mieter werden keine allgemeinen Versorgungsnetze verwendet. Darüber hinaus regelt der deutsche Gesetzgeber in diversen Gesetzen und Verordnungen weitere Detailpunkte des Mieterstroms. Neben dem Stromverkauf an Mieter kann der Vermieter nach wie vor den erzeugten Strom auch ins Stromnetz einspeisen und erhält dafür eine garantierte Einspeisevergütung. Diese liegt jedoch in der Regel um ca. 30 bis 50% niedriger als die Rückflüsse aus dem Direktverkauf an die Mieter. Aus diesem Grund muss für die ökonomische Rentabilität eine entsprechende Strommenge von den Mietern abgenommen werden, um die Mehrkosten für Smart Meter und Abrechnungssysteme zu kompensieren. Diese Mehrkosten sowie weitere Aspekte der Implementierung wurden daher bereits in ersten Studien für den deutschen Markt zusammengefasst [5], [14]. In Abbildung 1 ist ein typisches Verbrauchsprofil einer Immobilie mit vier Haushalten abgebildet, welche über eine Photovoltaikanlange (10kWp) sowie über eine Brennstoffzelle (3kW) verfügen. Es ist zu erkennen, dass nachts die Brennstoffzelle die Grundlast des Gebäudes vollständig deckt und die Photovoltaikanlage von ca. 4 bis 20 Uhr Strom produziert, allerdings nicht genug um den Peak am Abend abzudecken. Die dargestellten Verläufe sind hochgradig stochastisch, da sie von diversen Faktoren abhängen. So ist die Stromproduktion einer Photovoltaikanlange stark von Wettereinflüssen abhängig [15] und das Verbrauchsprofil vom Nutzerverhalten oder der Anwesenheit der Mieter [16]. Für ein Mieterstrommodell ergibt sich folglich eine doppelte Unsicherheit, bestehend aus der produzierten Strommenge sowie dem Strombedarf der Mieter.

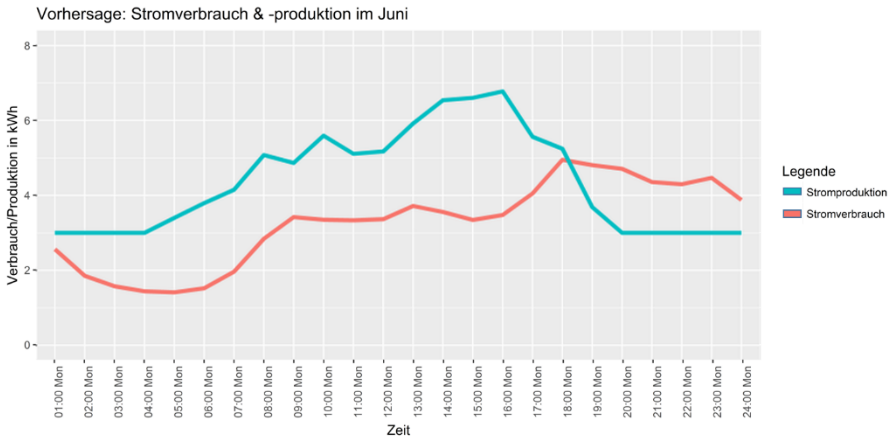


Abbildung 1. Beispielhafter Stromverbrauch und -produktion (3kW Brennstoffzelle und 10kWp Photovoltaikanlage) eines Mehrfamilienhauses (vier Einheiten)

Die für Mieterstrom eingesetzten Stromerzeugungstechniken Photovoltaik (PV) und BHKW weisen unterschiedliche Erzeugungsprofile auf. Die PV-Anlage besitzt sowohl einen Tages- als auch einen Jahresgang mit höchster Stromerzeugung zu der Mittagszeit und im Sommer. Dabei wird die Leistung einer PV-Anlage maßgeblich von der Umgebungstemperatur sowie der Globalstrahlung beeinflusst, und kann so zwischen einzelnen Tagen deutlich schwanken [15]. In der wissenschaftlichen Literatur ist daher in den vergangenen Jahren eine Vielzahl von Publikationen zur Vorhersage von Erzeugerprofilen bei PV-Anlagen entstanden. In der Regel werden dabei hybride Modelle verwendet, sodass auf Basis von Ingenieursgleichungen [17] lediglich die Eingangsvariablen der Gleichungen wie Temperatur- und Globalstrahlungsverläufe u.a. durch Neuronale Netze [18], Markov Modelle [19] oder Zeitreihen [20] stochastisch modelliert werden. Ein weiterer Teil der Literatur beschäftigt sich darüber hinaus mit Randparametern, wie zum Beispiel der Verschattung von PV-Anlagen [21] oder dem Anlagendesign [22]. Ein BHKW liefert dahingegen im durchgängigen Betrieb eine nahezu konstante Strommenge. Sollte die zusätzlich anfallende Wärmemenge im Sommer nicht benötigt werden, kann das Abschalten der Anlage rentabel sein. Die Berechnung des Eigenverbrauchs hat in den letzten Jahren durch den vermehrten Einsatz von Stromspeichern an Relevanz gewonnen [15], wobei zur Berechnung von Eigenverbräuchen in der Regel einzelne Objekte betrachtet oder Standardprofile verwendet werden.

Großklos [7] misst daher im Zuge der Investitionsbewertung von Mieterstrommodellen neben der Energieerzeugungsvorhersage auch dem Verbrauch und der tatsächlichen Stromabnahme eine wichtige Rolle bei. Da Stromverbrauchsprofile sowohl große Schwankungen im Tagesverlauf aufweisen als auch einem saisonalen Jahresverlauf unterliegen, sind für die Berechnung des Eigenverbrauchs möglichst hochauflösende Vorhersagen notwendig. Die verschiedenen Einflüsse auf Stromverbrauchsprofile untersuchen Sodenkamp et al. [9], indem sie mit Machine Learning Algorithmen und Verbrauchsdaten aus der Schweiz versuchen, anhand des Stromverbrauchs auf Eigenschaften eines Haushalts zu schließen (z.B. Alter des Hauses, die Verwendung von Wärmepumpen oder elektrischen Kochgeräten). In dieselbe Richtung arbeiten auch Beckel et al. [23], indem sie für einen irischen Datensatz diverse Eigenschaften zu Haushaltsbewohnern sowie des Gebäudes mit einer Trefferquote von zum Teil 70 bis 80% bestimmen. Für denselben Datensatz untersuchen McLouglin et al. [16] den Einfluss auf den Stromverbrauch von Faktoren, wie die Anzahl der Schlafzimmer, Zusammensetzung des Haushaltes oder den Wohnungstyp mit Hilfe einer multiplen linearen Regression. In einer späteren Arbeit wenden die Autoren dann statistische Clustermethoden wie k-Means, k-Medoids und Self Organising Maps an, um anhand des Stromverbrauches die Haushalte in verschiedene Klassen aufzuteilen [8]. Insgesamt identifizieren sie 10 Cluster, die jeweils charakteristische Verbrauchsverläufe aufweisen. Zu der gleichen Anzahl an Clustern kommen auch Haben et al. [24], unter Verwendung eines sogenannten Finite Mixture Clustering. Einen etwas anderen Weg schlagen Wang et al. [25] ein, indem sie erst die Verbrauchsdaten jedes Haushalts in ein Markov-Modell überführen und anschließend die Haushalte nach deren Wahrscheinlichkeitsdichten clustern, um so bestimmte Verhaltensweisen zu identifizieren. Abschließend schätzen Al-Wakeel et al. [26] fehlende Stromverbrauchswerte anhand des k-Means Algorithmus, indem sie verschiedene Distanzmaße von den Ausgangsdaten zu den Clusterzentren verwenden.

Die vorgestellten statistischen Methoden werden in der Praxis bei der ökonomischen Bewertung von Mieterstrom kaum angewandt, sodass bis heute Standardprofile oder Referenzprojekte zur Bewertung von Mieterstrommodellen herangezogen werden [7]. Ziel dieser Arbeit ist es daher, ein datenbasiertes Modell zur Vorhersage des haushaltsspezifischen Stromverbrauchprofils auf Basis weniger Faktoren, wie zum Beispiel der Anzahl der Zimmer, dem Beschäftigungsverhältnis des Mieters und der Haushaltsgröße zu entwickeln. Dieses Stromverbrauchsprofil kann anschließend zur Berechnung des Eigenverbrauchs verwendet werden, der wiederum als Grundlage für die Investitionsbewertung von Mieterstrommodellen dient.

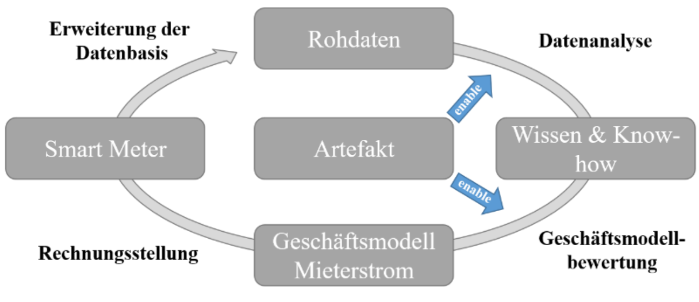


Abbildung 2. Datenkreislauf eines Geschäftsmodells zur Beratung bei Mieterstrom

Abbildung 2 zeigt das Artefakt dieses Beitrags entlang eines Datenkreislaufs für ein Geschäftsmodell zur Verwaltung und Beratung von Mieterstrom. Hierzu wird, ausgehend von Rohdaten zu Verbrauch und Haushaltseigenschaften, Know-how auf Grundlage von Datenanalysen aufgebaut. Dies ermöglicht im zweiten Schritt, in Form eines Bewertungsalgorithmus, die ökonomische Bewertung des Mieterstrommodells. Nach Umsetzung des Mieterstroms werden dann insbesondere zu Abrechnungszwecken weitere Daten über Smart Meter gesammelt, welche den Datenpool erweitern und so den gesamten Kreislauf kontinuierlich verbessern.

3 Modellteil

Im Folgenden werden die notwendigen Gleichungen zur Modellierung eines Mieterstrommodells auf Basis von PV und BHKW eingeführt. Um die ökonomische Bewertung eines Mieterstrommodells durchführen zu können, ist eine möglichst exakte Schätzung des Eigenverbrauchs notwendig. Auf dieser Basis lässt sich für ein Mieterstrommodell die folgende Gleichung des Cashflows im Jahr $t \in \{1, \ldots, T\}$ aufstellen, die Steuereffekte nicht berücksichtigt:

$$CF_t = E_t^G \cdot \varepsilon_t \cdot P_t^{MS} + E_t^G \cdot (1 - \varepsilon_t) \cdot P_t^{Netz} - K_t, \tag{1}$$

wobei $E_t^G \in \mathbb{R}^+$ die erzeugte Strommenge in kWh bezeichnet, $\varepsilon_t \in [0,1]$ den Eigenverbrauchsfaktor und P_t^{MS} den erzielten Preis pro kWh durch den Verkauf von Mieterstrom sowie P_t^{Netz} dementsprechend die Einspeisevergütung für die Stromabgabe ins Netz. Die Kosten K_t zum Zeitpunkt t setzen sich aus den Verwaltungskosten, den Abrechnungskosten (Smart Meter und IT-Infrastruktur) sowie ggf. anfallenden Kosten für Energieträger zusammen. Zur Vereinfachung werden die Erlöse aus der Wärmeerzeugung eines BHKW nicht berücksichtigt, können aber über verminderte Kosten modelliert werden. Hinter Gleichung (1) steckt die Annahme, dass zusätzlich benötigter Strom aus dem Stromnetz gewinnneutral an die Mieter weiterverkauft wird. Der Eigenverbrauchsfaktor ε_t lässt sich als Bruch von erzeugter Strommenge E_t^G und der von den Mietern abgenommenen Strommenge E_t^{MS} darstellen:

$$\varepsilon_t = \frac{E_t^{MS}}{E_t^G}. \tag{2}$$

Für eine gegebene Partition, welche einen Tag anhand von N Messpunkten in $N-1$ gleich große Intervalle teilt, kann die abgenommene Mieterstrommenge E_t^{MS} (auf Jahresbasis) wie folgt geschätzt werden:

$$E_t^{MS} \approx \sum_{i=1}^{365} \sum_{j=1}^{N} \min(e_{ij}, v_{ij}), \tag{3}$$

mit e_{ij} als die erzeugte Strommenge in kWh am Tag i zwischen Messpunkt j und $j-1$ sowie v_{ij} als der dazugehörige Stromverbrauch der Mieter. Dabei wird innerhalb der Summenzeichen der limitierende Faktor für den Mieterstrom berechnet, also die verfügbare Strommenge oder der Strombedarf der Mieter und dieser anschließend für alle Intervalle aufaddiert. Die gesamte erzeugte Strommenge E_t^G kann analog zu Gleichung (3) berechnet werden, indem lediglich die e_{ij} aufsummiert werden. Als nächster Schritt wird nun eine Methode zur Modellierung der erzeugten sowie verbrauchten Strommengen e_{ij} und v_{ij} eingeführt. Für die Modellierung der erzeugten Strommenge aus einer PV-Anlage e_{ij}^{PV} wird die Osterwald Methode [27] verwendet, welche trotz ihrer Einfachheit eine hohe Prognosegüte aufweist [17]. Demnach kann die Leistung L_{ij}^{PV} einer PV-Anlage über die folgende Gleichung definiert werden:

$$L_{ij}^{PV} = L_{ij}^{max}\left(G_{ij}, T_{ij}^{c}\right) \cdot (1 - L_V) \cdot (1 - L_R) \cdot (1 - L_T) \cdot (1 - L_S). \quad (4)$$

$L_V \in [0,1]$ bezeichnet die Verluste durch Verschattung, $L_R \in [0,1]$ definiert die Verluste durch Reflexion des Lichtes, $L_T \in [0,1]$ definiert die Verluste in den Leitungen des PV-Moduls beim Stromtransport sowie $L_S \in [0,1]$ die Systemverluste u.a. bei der Stromumwandlung. Die maximale Leistung $L_{ij}^{max}\left(G_{ij}, T_{ij}^{c}\right)$ in Abhängigkeit der Globalstrahlung G_{ij} und der Zellentemperatur der PV-Anlage T_{ij}^{c} ist definiert als:

$$L_{ij}^{max}\left(G_{ij}, T_{ij}^{c}\right) = L^{peak} \cdot G_{ij} \cdot \left[1 + 0{,}0035 \cdot \left(T_{ij}^{c} - T^{*}\right)\right] / G^{*}, \quad (5)$$

mit L^{peak} als die Gesamtleistung der PV-Anlage unter genormten Bedingungen bei einer Temperatur von $T^{*} = 25^{\circ}C$ und $G^{*} = 1000\, W/m^2$. Die Zellentemperatur T_{ij}^{c} einer PV-Anlage kann auf Basis der Außentemperatur T_{ij}^{A} sowie der Globalstrahlung G_{ij} wie folgt geschätzt werden [28]:

$$T_{ij}^{c} = T_{ij}^{A} + (T^{PV} - 20) + G_{ij}/800, \quad (6)$$

wobei T^{PV} die ausgelegte Zellentemperatur im Betriebsmodus ist. An dieser Stelle sei ergänzend vermerkt, dass die einfallende Globalstrahlung G_{ij} maßgeblich von der Ausrichtung der PV-Anlage und des Sonnenwinkels abhängt. Nach Reikard [20] sowie Fang und Lahdelma [29] können die Globalstrahlung G_{ij} und die Außentemperatur T_{ij}^{A} aufgrund ihrer statistischen Eigenschaft adäquat über Seasonal Autoregressive Integrated Moving Average (SARIMA) Modelle quantifiziert werden. Abschließend kann die Leistung L^{BHKW} eines BHKW vereinfachend als konstant angenommen werden, sodass sich die folgende Gleichung für die produzierte Strommenge e_{ij} ergibt:

$$e_{ij} = (L_{ij}^{PV} + L^{BHKW}) \cdot 24h/N. \quad (7)$$

Hinter Gleichung (7) steckt die Annahme, dass innerhalb eines Zeitschrittes die Leistungen der PV-Anlage wie auch des BHKWs konstant sind. Als zweiter Schritt wird nun die Modellierung des Stromverbrauchsprofils v_{ij} der am Mieterstrommodell beteiligten $K \in \mathbb{N}$ Haushalte eingeführt. Der Gesamtverbrauch v_{ij} lässt sich als die Summe der Einzelverbräuche der Haushalte darstellen:

$$v_{ij} = \sum_{k=1}^{K} v_{kij}, \quad (8)$$

wobei v_{kij} der Stromverbrauch des k-ten Haushaltes am i-ten Tag zwischen Messpunkt j und $j-1$ bezeichnet. Über jeden Haushalt sei ferner ein Merkmalsvektor $x \in \mathbb{N}^D$ bekannt, welcher aus $D \in \mathbb{N}$ einzelnen Merkmalen besteht. Zusätzlich sei bekannt, dass für Haushaltsstromverbräuche $Q \in \mathbb{N}$ Standardstromverbrauchsprofile existieren, anhand welcher sich Haushalte in verschiedene Cluster aufteilen lassen. Für jedes Cluster $q = 1, \dots Q$ seien $f_1, \dots, f_Q$ die Dichtefunktionen von $\mathbb{R}^D \to \mathbb{R}$, welche die Verteilung und Häufigkeit des Merkmalsvektors x in jedem Cluster charakterisieren. Zusätzlich seien die allgemeinen Wahrscheinlichkeiten $\pi_1, \dots, \pi_Q$ der Cluster bekannt, welche die Häufigkeiten der einzelnen Cluster charakterisieren und deshalb auch $\sum_{q=1}^{Q} \pi_q = 1$ gelten muss. Nach dem Satz von Bayes gilt dann für jeden Haushalt mit Merkmalsvektor x:

$$f_x(q) \coloneqq P(Haushalt\ gehört\ zu\ Cluster\ q\ |X = x) = \frac{\pi_q f_q(x)}{\sum_{l=1}^{Q} \pi_l f_l(x)}. \tag{9}$$

Mit Hilfe von Gleichung (9) kann somit die Wahrscheinlichkeit, dass ein Haushalt zu einem Cluster q gehört, anhand von verschiedenen Dichtefunktionen und Wahrscheinlichkeiten berechnet werden, die sich nach einer Clusteranalyse bestimmen lassen. Seien zum Beispiel $M \in \mathbb{N}$ Haushalte mit Merkmalsvektoren $x_1, \dots, x_M \in \mathbb{N}^D$, welche anhand ihres Stromverbrauchprofils in Q Cluster aufgeteilt wurden, gegeben, dann lassen sich die Schätzer $\hat{\pi}_q$ und $\hat{f}_q$ definieren durch:

$$\hat{\pi}_q = \frac{\#Anzahl\ Haushalte\ in\ Cluster\ q}{M}, \tag{10}$$

sowie

$$\hat{f}_q(x) = \frac{\#Anzahl\ Haushalte\ in\ Cluster\ q\ mit\ Merkmalsvektor\ x}{\#Anzahl\ Haushalte\ in\ Cluster\ q}. \tag{11}$$

Durch Einsetzen der beiden Schätzer in Gleichung (9) erhalten wir dann den kombinierten Schätzer:

$$\hat{f}_x(q) = \begin{cases} \dfrac{\hat{\pi}_q \hat{f}_q(x)}{\sum_{l=1}^{Q} \hat{\pi}_l \hat{f}_l(x)}, wenn \sum_{l=1}^{Q} \hat{f}_l(x) > 0 \\ \quad \hat{\pi}_q \qquad , sonst. \end{cases} \tag{12}$$

Der zweite Teil der Gleichung (12) dient der Robustheit des Schätzers, da der Nenner den Wert 0 annehmen würde, falls ein Merkmalsvektor x nicht in den ursprünglichen Daten enthalten wäre. Nun seien die durchschnittlichen Stromverbräuche innerhalb eines Clusters $\hat{v}_{ij}^q$ am Tag i zwischen Messpunkt j und $j-1$ bekannt für $q = 1, \dots, Q$. Dazu sei ergänzend bemerkt, dass die Verbräuche eines Clusters zu einer bestimmten Partition aus den historischen Verbräuchen zu den einzelnen Monaten, an einem durchschnittlichen Wochentag, über den betroffenen Zeitraum gemittelt werden können und so Schätzwerte für v_{ij}^q ableitbar sind. Daraus lässt sich der erwartete Verbrauch $\hat{v}_{kij}$ des k-ten Haushalts mit Merkmalsvektor x_k wie folgt berechnen:

$$\hat{v}_{kij} = \sum_{l=1}^{Q} \hat{f}_x(l) \cdot \hat{v}_{ij}^l. \tag{13}$$

Wird dieser Schätzer in die Ausgangsgleichung (8) eingesetzt, erhält man den entsprechenden Schätzer für den Stromverbrauch der Haushalte. Hinter dem eingeführten Merkmalsvektor x stehen numerisch codierte Merkmale, welche einen Haushalt charakterisieren und so auf den Verbrauch schließen lassen. Diese Merkmale, wie zum Beispiel die Anzahl der Haushaltsmitglieder, können sich über die Zeit ändern. Es wird jedoch zur Vereinfachung angenommen, dass dieser Vektor konstant ist.

Abschließend lässt sich nun der Kapitelwert (KW) des Mieterstrommodells mit Kalkulationszinssatz r und Investitionskosten I_0 zum Zeitpunkt $t = 0$ wie folgt berechnen:

$$KW = \sum_{t=1}^{T} \frac{CF_t}{(1+r)^t} - I_0. \tag{14}$$

Zudem ergibt sich die folgende Amortisationszeit (AZ):

$$AZ = \min_{t \in [1,T]} t + T \cdot \mathrm{I}_{\sum_{i=1}^{t} CF_i < I_o} \tag{15}$$

wobei $\mathrm{I} \in \{0,1\}$ definiert ist als Indikatorfunktion und somit AZ die Anzahl der Perioden definiert, bis die jährlichen Cashflows in Summe die anfänglichen Investitionskosten I_0 übersteigen, unter der Annahme das $AZ \leq T$.

4 Evaluation des Modells

In der folgenden Evaluation werden zwei Mieterstrominvestitionen auf Basis von Realweltdaten ökonomisch bewertet und versucht die Vorhersagegüte des Modells zu bestimmen [14]. Dabei wird ein Mieterstrommodell für vier Haushalte und eines für eine Quartierslösung (30 Haushalte) bewertet. Es wird angenommen, dass der Wärmebedarf der Beispielgebäude für die Heizung und Warmwasser nicht über Strom zur Verfügung gestellt wird. Zur Stromerzeugung werden bei den vier Haushalten eine $3kW$ Brennstoffzelle und eine $10kWp$ PV-Anlage installiert, welche annahmegemäß über 20 Jahre konstante Leistungen erbringen. In der Quartierlösung wird jeweils das 7,5-Fache dieser Leistung installiert. Damit liegt die Quartierslösung noch unter der vom Gesetzgeber vorgegeben Fördergrenze für Anlagen von insgesamt bis zu $100\ kW$. Zusätzlich wird über den Betrachtungszeitraum ein konstantes Nutzerverhalten der Mieter angenommen. Die Kostenersparnis durch Skaleneffekte bei der Quartierslösung entsprechen 10% pro kW-Leistung, wohingegen die Vergütung der Netzeinspeisung vom Fördergeber für Großprojekte, nach deutscher Förderrichtlinie, um 2 Cent reduziert wird. Weitere Annahmen und Parameter sind in Tabelle 1 zusammengefasst.

Tabelle 1. Modellparameter und Annahmen der Evaluation

Variante	T	r	I_0	K	P_t^{Netz}	P_t^{MS}	L_V	L_R	L_T	L_S	T^{PV}
4-Mieter	20	2%	50.000€	3.000€	0,12€	0,25€	0,16	0,07	0,00	0,12	47°C
Quartier			337.500€	20.250€	0,10€						

4.1 Vorstellung der Daten und Datenaufbereitung

Als Stromverbrauchsdaten werden öffentliche Smart Meter Daten von 4225 Haushalten der irischen Kommission für Energieregulierung (CER) verwendet, welche über einen Zeitraum von Juli 2009 bis Dezember 2010 alle 30 Minuten erfasst wurden [30]. Neben den Stromverbräuchen wurden diverse weitere Haushaltsmerkmale erfasst. Die verwendeten Merkmale für das Modell sind in Tabelle 2 zusammengefasst. Da die betrachteten Immobilien annahmegemäß über keine Stromheizung verfügen, werden sämtliche Haushalte aus den Daten entfernt, welche nicht dieser Annahme entsprechen oder deren Daten nicht vollständig sind. Danach werden für die vier meteorologischen Jahreszeiten, für einen durchschnittlichen Werktag, in jeweils dreistündigen Zeitintervallen ab 0 Uhr, der durchschnittliche Stromverbrauch je Haushalt sowie die Änderungsraten berechnet (64 Kennzahlen). Für die Modellierung der Temperatur und der Globalstrahlung (hor.) werden Daten vom Deutschen Wetterdienst für die Stadt Würzburg über den Zeitraum 2005 bis 2017 in stündlicher Auflösung verwendet [31].

Tabelle 2. Verwendete Merkmale zur Charakterisierung der Haushalte. In Klammern sind die Ausprägungen der vier Beispielhaushalte der ersten Investition dargestellt

Merkmale des Mieters	*Ausprägungsmöglichkeiten*
Alter Hauptmieter	<35$_{(4)}$, 36-45$_{(2)}$, 46-55$_{(3)}$, 56-65, >65$_{(1)}$
Beschäftigungsstatus	Angestellt$_{(2-4)}$, Selbstständig, Rente$_{(1)}$, Arbeitslos
Wohnform	Single$_{(1,4)}$, Eltern mit Kindern ≤ 15J.$_{(2)}$, Sonstige$_{(3)}$
Anzahl Bewohner > 15 Jahren	0$_{(1,4)}$, 1-2$_{(2,3)}$, >2
Anzahl Kinder ≤ 15 Jahren	0$_{(1,3,4)}$, 1$_{(2)}$, >1
Anzahl Schlafzimmer	1-2$_{(1,3,4)}$, 3$_{(2)}$, >3
Internetanschluss	vorhanden(v)$_{(2-4)}$, nicht vorhanden(n.v.)$_{(1)}$
Geschirrspülmaschine	vorhanden$_{(2-4)}$, nicht vorhanden$_{(1)}$
Spielekonsole	vorhanden, nicht vorhanden$_{(1-4)}$

4.2 Empirische Ergebnisse der Clusteranalyse

Zur Clusteranalyse wird der Partitioning Around Medoids (PAM) Algorithmus [32] und die Plattform R-CRAN sowie das Paket ‘cluster’ [33] genutzt. Die Auswahl der Clusteranzahl erfolgt über das Elbow-Kriterium, für welches ein Fehlermaß gegen die Anzahl der Cluster geplottet wird. Als Fehlermaß wird dabei die Summe der Entfernungsquadrate zwischen den Clustern zum Gruppenmittelpunkt des Clusters verwendet (Total within Sum of Squares). Mit zunehmender Anzahl der Cluster nimmt das Fehlermaß der Cluster ab, bis ab einem bestimmten Punkt eine deutliche Verflachung zu erkennen ist. Dieser Punkt markiert für die verwendeten Daten bei $Q = 6$ die optimale Anzahl der Cluster. Damit ist die gewählte Clusteranzahl etwas geringer als die vergleichbarer Arbeiten [8], was insbesondere darauf zurückzuführen ist, dass nur Haushalte betrachtet werden, die nicht mit Strom geheizt werden. In Abbildung 6 werden die durchschnittlichen Stromverbräuche an einem Werktag je Cluster dargestellt, worin sich die unterschiedlichen Verbrauchsprofile erkennen lassen. In Tabelle 3 sind die Ausprägungen dargestellt, welche im Vergleich zum Durchschnitt besonders häufig in den Clustern auftreten. Daraus lassen sich grob ein Cluster für Singles (1), drei Cluster für Familien in den unterschiedlichsten Lebensphasen (2-4), ein Cluster für Paare (5) und eines für Rentner (6) ableiten. Zur Modellierung der SARIMA-Prozesse für den Verlauf der Globalstrahlung und der Temperaturkurven wird die Plattform R CRAN und das Paket 'forecast' verwendet [34]. Dabei werden für jeden Kalendermonat auf stündlicher Basis, für die Stunden zwischen Sonnenaufgang und -untergang mit Dauer $s_i \in \mathrm{N}$, für $i = 1, ..12$ jeweils zwei SARIMA$(p, d, q) \times (P, D, Q)_{s_i}$ Modelle für $p, d, q, P, D, Q \leq 1$, gefittet und nach dem AIC-Kriterium ausgewählt, sodass sich 24 Zeitreihen ergeben. Um die Modellgüte der Zeitreihen als gegeben beurteilen zu können, werden im Anschluss die Modellparameter auf Signifikanz ($\alpha = 0.05$) untersucht, die Residuen mit Hilfe des Ljung–Box Tests auf Autokorrelation getestet sowie auf die Standardnormalverteilung untersucht.

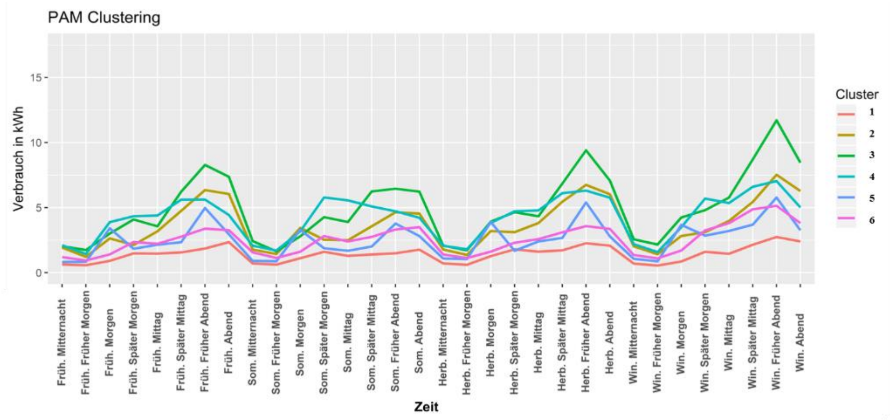


Abbildung 3. Stromverbräuche der Cluster nach Jahres- und Tageszeiten (Werktag) auf Basis von Smart Meter Daten der irischen Kommission für Energieregulierung (CER) [30]

Tabelle 3. Darstellung der Merkmale je Cluster die um 10 Prozentpunkte gestiegen sind zum Durchschnitt (~) auf Basis des Clusterings von Stromverbrauchsdaten [30]

Merkmale /Cluster	*1(rot)*	*2(braun)*	*3(grün)*	*4(h.blau)*	*5(blau)*	*6(pink)*
Alter des Hauptmieters	~	36-55	46-55	~	~	>65
Beschäftigungsstatus	~	Angest.	Selbst.	~	Angest.	Rentner
Wohnform	Single	Familie	Familie	Sonstige	~	Sonstige
Anzahl Mitbewohner über 15 Jahren	0	1-2	>2	>2	1-2	~
Anzahl Kinder unter 15 Jahren	0	1,>1	>1	~	~	0
Anzahl Schlafzimmer	1-2, 3	>3	>3	>3	~	~
Internetanschluss	n.v.	v.	v.	v.	~	~
Geschirrspülmaschine	n.v.	v.	v.	v.	~	~
Spielekonsolen	n.v.	v.	v.	v.	~	n.v.
Anteil Haushalte (%)	27,3	15,4	9,1	9,6	11,8	26,8

4.3 Monte Carlo Simulation und empirische Ergebnisse

Auf Basis der in Abschnitt 4.2 kalibrierten Modelle werden nun für vier bzw. dreißig zufällig ausgewählte Haushalte aus dem Datensatz die in Tabelle 1 beschriebenen Mieterstrommodelle ökonomisch und ökologisch bewertet. Ziel ist es, die Erwartungswerte des Eigenverbrauchsfaktors ε_t sowie den der erzeugten Strommenge E_t^G auf Basis einer Monte Carlo Simulation mit 1.000 Durchläufen vorherzusagen. Auf dieser Basis können dann der Kapitalwert und die Amortisationszeit der Investitionen berechnet werden. Mit Erwartungswerten in Höhe von 33.400 kWh/a für die erzeugte Strommenge sowie von 65% für den Eigenverbrauchsfaktor, ergeben sich bei dem Mieterstrommodell mit vier Haushalten eine Amortisationszeit von 13 Jahren sowie ein erwarteter Kapitalwert von ca. 12.000€ über den Betrachtungszeitraum. Dem gegenüber steht die Quartierslösung mit einer erwarteten Amortisationszeit von 10 Jahren sowie einem erwarteten Kapitelwert von ca. 220.000€. Trotz der geringeren Einspeisevergütung amortisiert sich die Quartierslösung somit im Schnitt drei Jahre vor dem Mieterstrommodell mit nur vier Haushalten. Dies liegt unter anderem an dem höheren Eigenverbrauchsfaktor der Quartierslösung von insgesamt 76%, welcher in Tabelle 4 auf monatlicher Basis dargestellt ist. Der höhere Eigenverbrauchsfaktor lässt sich über die rechtsschiefe Verteilung der Stromverbräuche erklären, sodass mit einer steigenden Anzahl an Mietern, Haushalte mit hohen Verbräuchen statistisch häufiger auftreten und somit auch mehr Strom abgenommen wird. Insgesamt wird auf Grundlage des Anlagendesigns die 7,5-Fache Strommenge im Quartier produziert. Würden sich die Investitionskosten I_0 und die jährlichen Kosten K der Quartierslösung im Vergleich zum Mieterstrommodell mit vier Haushalte um jeweils 20% verringern pro kW-Leistung (davor 10%), wird eine Amortisationszeit von 8 Jahren erreicht. Diese Einsparungen können in etwa als obere Grenze für die betrachtete Quartierslösung gelten [35]. Wie in Tabelle 4 dargestellt, tritt bei der Vorhersage des monatlichen Eigenverbrauchsfaktors bei der Investition mit vier Mietern ein absoluter Fehler von 3,5 Prozentpunkten und bei der Quartierslösung von 2,5 Prozentpunkten auf.

Table 4: Empirische Ergebnisse des Entscheidungsunterstützungsmodells

4-Mieter	Jan.	Feb.	Mär.	Apr.	Mai	Jun.	Jul.	Aug.	Sep.	Okt.	Nov.	Dez.
$E[\varepsilon]$	0,79	0,72	0,65	0,59	0,56	0,55	0,57	0,58	0,62	0,67	0,74	0,84
ε_{real}	0,75	0,69	0,60	0,58	0,54	0,53	0,55	0,55	0,58	0,63	0,69	0,77
Quartier												
$E[\varepsilon]$	0,88	0,83	0,77	0,70	0,69	0,67	0,69	0,70	0,74	0,79	0,84	0,90
ε_{real}	0,86	0,83	0,77	0,65	0,63	0,67	0,67	0,67	0,70	0,78	0,82	0,87

Wird unter Betrachtung des gesamten Lebenszyklus der Mieterstrominvestition die CO_2-Einsparung berechnet, ergeben sich für die Quartierslösung eine jährliche Einsparung in Höhe von 95 Tonnen CO_2, was einer Reduktion von 64% im Vergleich zum deutschen Kraftwerksmix gleichkommt.[1] Dies entspricht in etwa dem CO_2-

[1] Es wurde ein CO_2-Ausstoß pro kWh Strom von 100 Gramm bei PV, 240 Gramm bei einer Brennstoffzelle und 550 Gramm im deutschen Kraftwerksmix angenommen [36].

Ausstoß eines Mittelklassewagens über ca. 600.000 km Fahrstrecke. Für das zweite Szenario ergibt sich eine jährliche Einsparung von 13 Tonnen CO_2. Der durchschnittliche Stromverbrauch pro irischem Haushalt von 3.700 kWh/a, auf dessen Basis die Ergebnisse berechnet wurden, liegt leicht über dem durchschnittlichen Verbrauch von 3.200 kWh/a in Deutschland [37]. Ein auf dieser Grundlage kalibriertes Modell wird den Strombedarf deutscher Haushalte überschätzen und dadurch tendenziell zu kurze Amortisationszeiten berechnen. Wird der durchschnittliche Mehrverbrauch der irischen Haushalte pauschal abgezogen, ergeben sich für Deutschland im Schnitt ein Jahr längere Amortisationszeiten.

5 Zusammenfassung, Limitationen und Ausblick

Für den Erfolg der Energiewende spielt die dezentrale Stromerzeugung eine entscheidende Rolle. Dabei bietet Mieterstrom für die auftretenden Herausforderungen einen vielversprechenden Lösungsansatz, welcher insbesondere durch die Nutzung des steigenden Datenvolumens durch Smart Meter zu neuen Geschäftsmodellen führen kann. In dieser Arbeit wurde deshalb zur Investitionsbewertung von Mieterstrommodellen ein Entscheidungsunterstützungssystem entwickelt und evaluiert. Die Ergebnisse der Wirtschaftlichkeitsbetrachtung zeigen, dass sich Mieterstrommodelle unter den getroffenen Annahmen je nach Größe nach 8 bis 14 Jahren amortisieren. Zusätzlich konnten bei der hybriden Nutzung von PV-Anlagen und Brennstoffzellen $C0_2$-Einsparungen von über 60% nachgewiesen werden. Zu den Defiziten der Annahmen: (1) Ein gleichbleibendes Nutzverhalten kann nur schwer über lange Zeiträume angenommen werden, da die natürliche Veränderung der Lebenssituation eines Haushaltes zu Veränderungen im Verbrauchsverhalten führen kann. Diese Annahme könnte jedoch mit einem zusätzlichen Teilmodell zur Fortschreibung der Haushaltsmerkmale entkräftet werden. (2) Es werden keine Ausfälle durch Vertragskündigungen berücksichtigt, wie sie zum Beispiel durch Mieterwechsel entstehen. An dieser Stelle könnten Langzeitstudien den notwendigen Aufschluss über die Häufigkeit dieser Ausfälle liefern. (3) Die berechneten Amortisationszeiten sollten auf Basis deutscher Verbrauchsdaten validiert werden, um so die vollständige Übertragbarkeit auf Deutschland zu gewährleisten. (4) Entstehende Preisrisiken durch die stochastische Preisentwicklung für Energieträger und die zukünftige Entwicklung der Betriebskosten des Mieterstrommodells werden aktuell noch nicht abgebildet. Dies kann jedoch auf Basis von bestehender Literatur ergänzt werden. Zukünftige Erweiterungen des Modells sollten neben den vier genannten Limitationen insbesondere die Integration von Speichertechnologien sowie das Zusammenspiel von Mieterstrom und Elektrofahrzeugen berücksichtigen. Durch diese Trends könnten sich die Profile von abgerufenen und bereitgestellten Strommengen signifikant ändern und so zu neuen Rahmenbedingungen für Mieterstrommodelle führen. Um ein beratungs- und serviceorientiertes Geschäftsmodell basierend auf dem entwickelten Ansatz zu realisieren, ist zusätzlich eine allumfassende IT-Unterstützung notwendig. Hier ist es weiterhin notwendig die technische Ausgestaltung solcher IT-Anwendungen im Detail zu betrachten. In diesem Zusammenhang spielt auch das

Thema Datenschutz eine wichtige Rollte, da die Privatsphäre der Mieter ebenfalls geschützt werden muss. Trotz der genannten Limitation zeigt der vorliegende Beitrag auf, wie ein Entscheidungsunterstützungssystem für Mieterstrom konzipiert sein müsste und kann somit zukünftig zur Identifikation und Umsetzung erfolgreicher Investitionen beitragen.

Literatur

1. Davis, S.J., Caldeira, K., Matthews, H.D.: Future CO2 emissions and climate change from existing energy infrastructure. Science (New York, N.Y.) 329, 1330–1333 (2010)
2. Hake, J.-F., Fischer, W., Venghaus, S., Weckenbrock, C.: The German Energiewende – History and status quo. Energy 92, 532–546 (2015)
3. Pepermans, G., Driesen, J., Haeseldonckx, D., Belmans, R., D'haeseleer, W.: Distributed generation: definition, benefits and issues. Energy Policy 33, 787–798 (2005)
4. Behr, I., Großklos, M.: Mieterstrom – ein Beitrag zur Energiewende. In: Behr, I., Großklos, M. (eds.) Praxishandbuch Mieterstrom, pp. 3–14. Springer Fachmedien, Wiesbaden (2017)
5. Beucker, S., Hinterholzer, S., Schweinoch, M., Zernahle, O.: Messkonzepte für Mieterstrom. In: Behr, I., Großklos, M. (eds.) Praxishandbuch Mieterstrom, 36, pp. 81–98. Springer Fachmedien, Wiesbaden (2017)
6. Leprich, U., Neumann, W.: Zur Rolle des Mieterstroms in Energiesystemen der Zukunft. In: Behr, I., Großklos, M. (eds.) Praxishandbuch Mieterstrom, pp. 15–27. Springer Fachmedien, Wiesbaden (2017)
7. Großklos, M.: Erzeugungs- und Verbrauchsprofile für die Planung von Mieterstrom. In: Behr, I., Großklos, M. (eds.) Praxishandbuch Mieterstrom, pp. 113–123. Springer Fachmedien, Wiesbaden (2017)
8. McLoughlin, F., Duffy, A., Conlon, M.: A clustering approach to domestic electricity load profile characterisation using smart metering data. Applied Energy 141, 190–199 (2015)
9. Sodenkamp, M., Kozlovskiy, I., Hopf, K., Staake, T.: Smart Meter Data Analytics for Enhanced Energy Efficiency in the Residential Sector. In: Wirtschaftsinformatik 2017 Proceedings. AIS electronic library, St. Gallen. (2017)
10. Hopf, K., Sodenkamp, M., Kozlovkiy, I., Staake, T.: Feature extraction and filtering for household classification based on smart electricity meter data. Comput Sci Res Dev 31, 141–148 (2016)
11. Watson, R.T., Boudreau, M.-C., Chen, A.J.: Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions for the IS Community. MIS Quarterly 34, 23 (2010)
12. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75 (2004)
13. Nestle, U., Kunz, C.: Studienvergleich: Stromgestehungskosten verschiedener Erzeugungstechnologien. Agentur für erneuerbare Energien, Berlin, (2014)
14. Paschka, D.: Wirtschaftlichkeit von Mieterstrom. In: Behr, I., Großklos, M. (eds.) Praxishandbuch Mieterstrom, pp. 125–139. Springer Fachmedien, Wiesbaden (2017)
15. Masa-Bote, D., Castillo-Cagigal, M., Matallanas, E., Caamaño-Martín, E., Gutiérrez, A., Monasterio-Huelín, F., Jiménez-Leube, J.: Improving photovoltaics grid integration through short time forecasting and self-consumption. Applied Energy 125, 103–113 (2014)
16. McLoughlin, F., Duffy, A., Conlon, M.: Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. Energy and Buildings 48, 240–248 (2012)

17. Fuentes, M., Nofuentes, G., Aguilera, J., Talavera, D.L., Castro, M.: Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in Mediterranean climates. Solar Energy 81, 1396–1408 (2007)
18. Mellit, A., Pavan, A.M.: A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy 84, 807–821 (2010)
19. Mellit, A., Benghanem, M., Arab, A.H., Guessoum, A.: A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. Solar Energy 79, 469–482 (2005)
20. Reikard, G.: Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy 83, 342–349 (2009)
21. Masa-Bote, D., Caamaño-Martín, E.: Methodology for estimating building integrated photovoltaics electricity production under shadowing conditions and case study. Renewable and Sustainable Energy Reviews 31, 492–500 (2014)
22. Yin, E., Li, Q., Xuan, Y.: Optimal design method for concentrating photovoltaic-thermoelectric hybrid system. Applied Energy 226, 320–329 (2018)
23. Beckel, C., Sadamori, L., Staake, T., Santini, S.: Revealing household characteristics from smart meter data. Energy 78, 397–410 (2014)
24. Haben, S., Singleton, C., Grindrod, P.: Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Data. IEEE Trans. Smart Grid 7, 136–144 (2016)
25. Wang, Y., Chen, Q., Kang, C., Xia, Q.: Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications. IEEE Trans. Smart Grid 7, 2437–2447 (2016)
26. Al-Wakeel, A., Wu, J., Jenkins, N.: k -means based load estimation of domestic smart meter measurements. Applied Energy 194, 333–342 (2017)
27. Osterwald C.R.: Translation of device performance measurements to reference conditions. Solar Cells, 269–279. (1986)
28. International Electrotechnical Commission: Crystalline silicon terrestrial photovoltaic (PV) modules—design qualification and type approval. International Standard IEC 61215 (2005)
29. Fang, T., Lahdelma, R.: Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating. Applied Energy 179, 544–552 (2016)
30. Commission for Energy Regulation (CER): CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition, www.ucd.ie/issda/CER-electricity (Accessed: 12.09.2018)
31. DWD Climate Data Center: Historische stündliche Stationsmessungen für Deutschland, ftp://ftp-cdc.dwd.de/pub/CDC/(Accessed: 12.09.2018)
32. Kaufman, L., Rousseeuw, P.J.: Finding Groups in Data. An Introduction to Cluster Analysis. John Wiley & Sons Inc, Hoboken (2009)
33. Maechler, M., Rousseeuw, P., Struyf, A.: cluster - Cluster Analysis Basics and Extensions. R package version 2.0.7-1 (2018)
34. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, 1-22 (2008)
35. BKI-Baukosteninformationszentrum Deutscher Architektenkammern, B.-B.D.: BKI Baupreise kompakt 2018 - Neu- und Altbau. Statistische Baupreise für Positionen mit Kurztexten. BKI, Stuttgart (2017)
36. Memmler, M., Lauf, T., Wolf, K., Schneider S. (eds.): Emissionsbilanz erneuerbarer Energieträger. Bestimmung der vermiedenen Emissionen im Jahr 2016. Im Auftrag des Umweltbundesamtes (2017)
37. Sustainable Energy Authority of Ireland: Energy in the residential sector. 2018 Report (2018)

Discovering Blockchain for Sustainable Product-Service Systems to enhance the Circular Economy

Jannis Vogel[1], Simon Hagen[1] and Oliver Thomas[1]

[1] University of Osnabrueck, Chair of Information Management and Information Systems
Katharinenstr. 3, 49074 Osnabrueck, Germany
{jannis.vogel,simon.hagen,oliver.thomas}@uni-osnabrueck.de

Abstract. An increasing amount of use cases is discovered for blockchain technology, since it promises tamper-proof recording of product-related data. It has the potential to improve the reliability of information management for whole supply chains and thus enables new ecologically and economically service offerings. Integrating products and services into one marketable bundle is no new concept and is referred to as product-service systems (PSS). Therefore, the methodical integration of knowledge on sustainable businesses, PSS and blockchain is a promising approach to overcome current barriers to achieve an applicable circular economy. Our study contributes a structured literature review on ongoing research in the field of sustainability-focused blockchain applications. From this, we elaborate a holistic perspective by the integration of key concepts from two additional literature reviews for blockchain and PSS. As a result, we point out potential benefits and present the effect of blockchain on sustainable PSS with a product-life cycle model.

Keywords: Blockchain, Product-Service Systems, Circular Economy, Supply Chain

1 Introduction and Motivation

Circular Economy is a concept to increase the efficiency of resources and to implement sustainability in product-life cycles that will interrupt the phase of pure resource extraction, production, usage, and waste [1]. Barriers such as the customer's perception, information management, and missing performance indicators hinder a comprehensive application in businesses [2]. Hence new approaches are necessary to solve current problems in the domain of circular economy. One promising approach is the technology Blockchain. Blockchain has evolved as a capable technology to enhance information exchanges between individuals [3]. This contribution investigates how blockchain technology can improve sustainable business activities in particular regarding the circular economy.

However, in Gartner's Hype Cycle of uprising technologies, the blockchain is currently in the "trough of disillusion", expectations were thus not met [4]. From an environmental perspective, the blockchain is even more in a negative focus since it has a high demand for resources (e.g. for mining the cryptocurrency Bitcoin) [5]. The new

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

technology raises many unanswered research questions, e.g. how blockchain systems have to be designed to be sustainable itself and follow concepts like Green by IT. In contrast, a disseminated research approach is the design of product-service systems (PSS) that have an evidenced impact on sustainability [6].

Hence, we propose the integration of blockchain as a technology, i.e. as enabler, and PSS as a business strategy to create ecologic and economic sustainable offerings. Blockchain allows the independent and tamper-proof record of product-related data and the traceable accountability regarding reuse or recycling [7]. Furthermore, the enhanced transparency of product components and materials they are made of contributes to a simplified recycling process and the estimation of resource values [8]. To obtain a holistic view on the concerned fields (blockchain, PSS and sustainability, i.e. circular economy) we conducted a literature review to identify current sustainable-, product- and service-related blockchain-based use cases. In a next step, we combined our research results with blockchain [3] and sustainable PSS characteristics [6] to determine which blockchain-based use cases subject to precise blockchain characteristics and how they influence existing sustainable PSS characteristics. By doing so, we answer the following research questions:

1.) *What are current blockchain-based use cases and how do they influence sustainable PSS characteristics?*
2.) *How do blockchain-based sustainable PSS influence the use of resources and enhance the circular economy?*

By the combination, we designed a blockchain-based product-life cycle model for PSS according to Blinn et al. [9]. Thereby the model demonstrates what kind of impact the application of blockchain in PSS has onto the product-life and resource efficiency. Overall the paper contributes first proposals how blockchain and PSS can be combined to achieve sustainable potentials in the circular economy. The blockchain-based product-life cycle model can be used to redesign PSS with the application of blockchain in a more sustainable manner and to overcome barriers in the circular economy. The remainder of this article is structured as follows: Section 2 explains the applied method to streamline this research. Followed by section 3 that describes the fundamentals regarding blockchain and product-service systems. Section 4 presents the findings based on the conducted literature research and the combination of blockchain and PSS characteristics. Section 5 presents based on the literature review the blockchain-based product-life cycle model. Finally, section 6 discusses limitations and further research needs.

2 Method

In order to clarify the research questions stated above, we conducted a literature review [10] to obtain an overview of current use cases regarding the application of blockchain technology in product, service and sustainability domains. The applied literature search process (cf. figure 1) is based on Dybå and Dingsøyr [11]. We queried the scientific databases *AISeL*, *Google Scholar*, *ScienceDirect*, *SpringerLink* and *Web of Science* with the term "Blockchain" AND ("Circular Economy" OR "Supply Chain" OR "Sustainability" OR "Environment" OR "Recycling" OR "Provenance" OR "WEEE" OR "Service") and their German translations. No limitations regarding the year of publication or any other attribute were made. Initially, this brought up 2393 sources in total, from which we removed articles by title and abstract. The process revealed 51 relevant articles that were then used for an in-depth analysis. In doing so, we focused on practical use cases of blockchain technology for sustainability with particular importance on saving data of a product during its complete lifecycle, cp. Blinn et al. [9]. The findings of the remaining 22 publications were then transferred into a concept matrix according to Webster and Watson [12] (cf. figure 3).

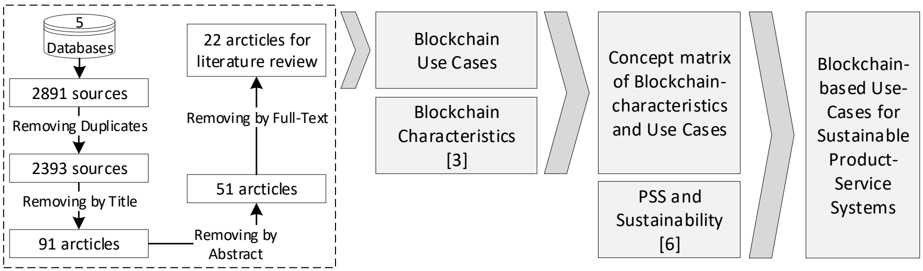


Figure 1. Literature Research Process and derived Research Outcome

The identified blockchain use cases were then combined with the blockchain characteristics examined in a literature review by Seebacher and Schüritz [3], in order to classify which blockchain-related characteristics realize the identified use cases. In a next step, the combined matrix is compared to factors of sustainability of PSS, which were analyzed by Hüer et al. [6], who also performed a literature review. The combination enables us to link all three literature reviews in an intertwined matrix (cf. figure 3) to answer what kind of blockchain-based use cases are realized by which blockchain characteristics and how they influence sustainable PSS characteristics. The related impacts on a circular economy are examined with a product-life cycle model for blockchain-based PSS (cf. figure 4) in accordance to Blinn et al. [9].

3 Fundamentals

3.1 Blockchain

We follow the Blockchain definition by Seebacher and Schürlitz for this paper [3]: “A blockchain is a distributed database, which is shared among and agreed upon a peer-to-peer network. It consists of a linked sequence of blocks, holding timestamped transactions that are secured by public-key cryptography and verified by the network community. Once an element is appended to the blockchain, it cannot be altered, turning a blockchain into an immutable record of past activity.” Thereby blockchain fundamentals are based on the whitepaper by Nakamoto [13] in 2008. In the last decade, different application areas evolved as postulated by Nofer et al. [14]. Li et al. [15] also investigate current blockchain applications within the business context with a literature review. They discovered that most blockchain-based articles focus on how the blockchain technology works and neglect to answer on what and why in terms of potential use cases and motivation factors that make blockchain favorable. Further related work that identified business application areas were contributed by Konstantinidis et al. [16] and by Witt and Richter [17] showing a high interest in practice and academic. Finally, Risius and Spohrer [18] developed a research framework for blockchain systems to guide researches and practitioners.

3.2 Product-Service Systems

PSS follow the idea of combining physical products and intangible services into one marketable bundle, which in combination serves the demand of the user [19, 20] and have a holistic view on the life-cycle of an offering, realized in an "extended value creation network" [21]. This is to be achieved by providing the right amount of product and service share (cf. figure 2) and therefore generating value for the customer, not by onetime selling a product or service solely. An example how this can be achieved is by providing the "solution" for a certain period, which leads to constant revenue for the offering company and calculable value and costs for the customer [22], e.g. by leasing a machine. Morelli [23] defines a PSS from three perspectives: First, from a traditional marketing point of view, where an entity can be reduced to its material components, to "an entity whose material component is inseparable from an immaterial one". Second, from a marketing service perspective, which shifts from "standardized services towards personalized ones". Lastly, from a product management perspective, the substitution of (physical) product shares by services and vice versa.

According to this, the "bundle" is not required to have a certain product or service share [24] and can be offered by a single company or a company-network [25]. The different shares of the components within a PSS are shown in figure 2. As one can see, the transition is fluent and Tukker [26] groups the different partitions into three core types, namely product, use, and result-oriented PSS.

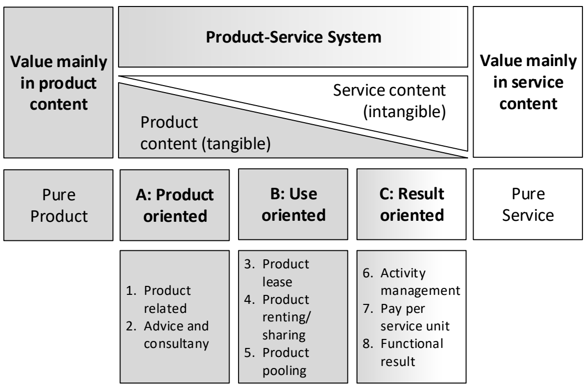


Figure 2. Product-Service System Classification according to Tukker [26]

Since PSS have a holistic view of the product itself and its life-cycle, many scientific publications examine the potentials of PSS regarding sustainability, e.g. by Tukker [26]. However, even though different PSS enabled business models, like renting or leasing, seem to have a positive influence on sustainability in the first instance, PSS are not a "sustainability panacea" [27]. Hüer et al. [6] performed a literature review on how PSS influence the sustainability and found six distinct characteristics: Substitution, Ownership, Product-Life, Cooperation, Sharing, Consumption. Though, the characteristics identified have both positive and negative influences on sustainability. Also, only a few practical methods on how to develop or use PSS in a sustainable way were proposed [6]. This circumstance is, amongst others, due to a perceived reduced user experience of PSS compared to "regular" products and the very different skills required to implement PSS. Thus, they have the potential to enable resource efficiency and ecologic sustainability, but key drivers, e.g. technologies or user experience, are missing for wide implementation.

4 Literature Review on Blockchain-based Use Cases

4.1 Classification of Blockchain-based Use Cases

We identified 22 relevant papers in our literature study, which revealed four closely related categories during their analysis. The first category contains literature that addresses the aforementioned research domain directly, referring to blockchain-based applications in the circular economy and sustainability domain. The second category consists of blockchain-based literature that focuses mainly on products while the third category relates to service and sharing economy aspects. The fourth category includes blockchain-based applications that support a tamper-proof track and traceability of supply chains to collect additional information that can lead to new services. An overview of the identified use cases, how they are realized and what kind of influence they have on PSS is shown in figure 3 and addresses the first research question.

In total 14 blockchain-based use cases are identified in the literature review which are explained in more detail in the following sections (4.2 – 4.5). Four blockchain characteristics mainly implement the identified use cases: Trust, shared and public, immutability, and decentralization. According to Seebacher and Schürlitz [3] trust is achieved amongst other blockchain characteristics principal through the tamper-proof data recording and the shared and public data transmission in a network. Decentralization is supported by the characteristics stated above and allow the creation of a network. Furthermore, decentralization enables trust with the consensus mechanism that removes the need for a trusted third-party [3].

The identified blockchain use cases have an impact on existing PSS characteristics. The concept of *substitution*, i.e. replacing products with customer-oriented services [6], is affected by new blockchain-based concepts for a sharing economy [28], [29] that allow a higher customer acceptance. Furthermore, blockchain influences the *ownership* concept. The PSS characteristic assumes that a product is designed more sustainable if the ownership belongs to the producers since they have direct incentives to create a robust product, perform maintenances and to design more recyclable products [6]. We expect a decreased influence of ownership in a sustainable PSS concept. This is justified with blockchain use cases such as product stewardship [7], [30], [31], [32], product history [30], [31], [33–36] and track and tracing systems [31], [34], [37–40] that allow a constant product traceability and a shared disclosure of product quality ensuring a permanent incentive to generate sustainable PSS without the importance of ownership. *Sharing* activities can enhance the utilization of products [6] due to simplified reselling [30], [32], [35], [41], [42] and sharing of goods [28], [29] with blockchain based on continuous tracking of the product's life and value that increases sharing activities. The *product life* is influenced by collaboration [6] and has a close relationship with the *ownership* concept. For instance, blockchain can influence collaboration with increased efficiency in the overhaul process of products with a more effortlessly access to maintenance-related data [31] or reduces adverse selection [42] and facilitate collaboration regarding the sales process. Moreover, blockchain has an impact on *consumption* by changing the customer's awareness for product quality and sustainability. Through a transparent supply chain e.g. corporate social responsibility activities can be tracked and communicated more reliably [33], [41]. Additionally, new services that are linked with the product using new information exchange and coordination mechanisms can be created [3], [43]. Overall the explicit information about the product's origin and quality lead to a higher willingness-to-pay for excellent quality products [36]. Blockchain-based supply chain management [7], [31], [40], [44–46] leads to a transparent tamper-proof product history. This can result in increased *cooperation* activities because of higher trust and shared information without the need of a third-party. In terms of PSS – cooperation has an overall positive effect on sustainability [6].

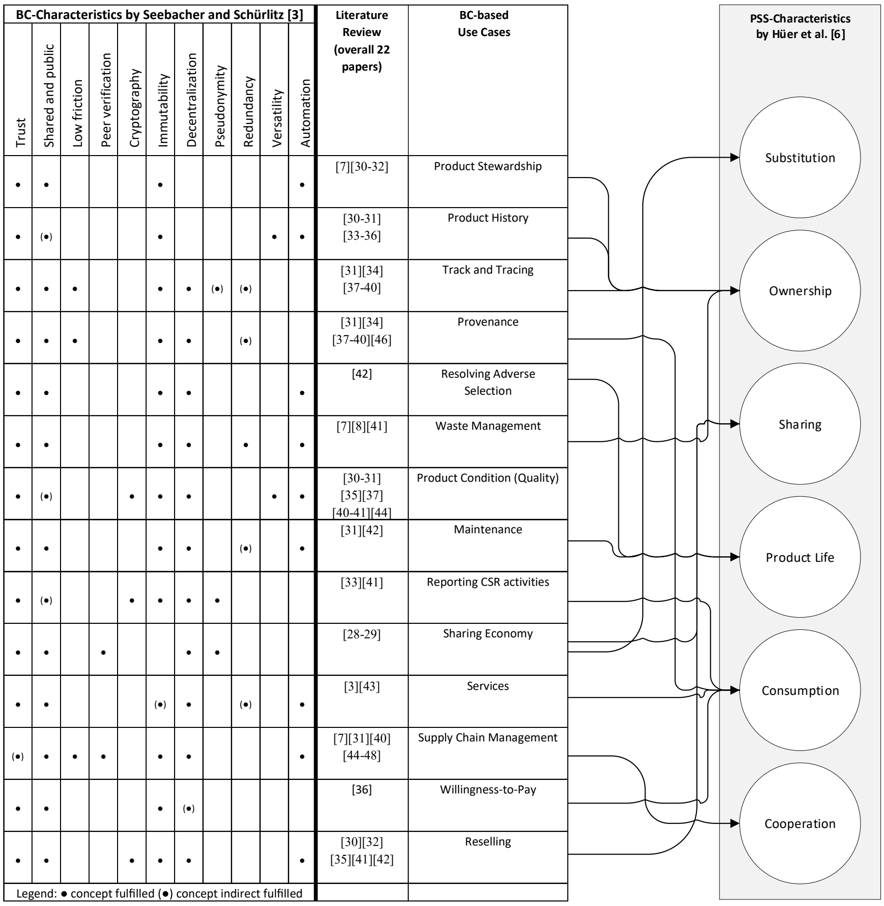


Figure 3. BC-based Use Cases realized by BC-Characteristics influencing PSS-Characteristics

4.2 Sustainability-related Blockchain-based Use Cases

Saberi et al. [7] propose a sustainable supply chain management concept that applies to electronic components which have to accomplish the Waste Electrical and Electronic Equipment Directive (WEEE Directive). They suggest blockchain as a feasible technology to effect higher product stewardship that positively affects the circular economy. In their point of view, the main benefits of blockchain are an easily accessible product history for every stakeholder in the supply chain, the tamper-proof storing of information and the resulting trust, smart contracts that enable to trace products and Internet of Things (IoT) devices that deliver reliable information gathered from the product or during processes. Ongena et al. [8] have a similar approach by working on waste management in general. They implemented a blockchain-based prototype and investigated if blockchain is a suitable technology to solve current issues in waste management. They discovered that a blockchain-based system for waste management addresses the deficits of information losses with the help of digitalization. However, they also found that blockchain-based systems solely cannot solve fraud and manipulations in waste management, due to the lack of control over the correctness of source data from other actors in the supply chain.

This shortcoming is tackled by Askoxylakis et al. [30] by using IoT devices to collect the "location, condition and availability" of assets to realize a blockchain-based system to exchange and collect the condition of assets. They propose to use these three properties to gather the value of specific assets. In addition, they suggest saving additional information that are relevant for customers into the blockchain, such as defects, maintenances or changes. In their opinion blockchain and IoT are the key drivers to enhance the circular economy with the gained data. Düring and Fishbeck [41] propose to use the blockchain and IoT components to create transparency in supply chains. Thereby, they especially mention the processes after a product is sold such as usage, reselling and recycling. Manufacturers could gain highly useful information to enhance their products with this kind of approach.

Due to e.g. the European Directive 2014/95/EU large enterprises are obligated to track their sustainability activities, also known as corporate social responsibilities (CSR), and publish an annual report [47]. Given that smaller companies are the contractor of obligated firms and perform services within the supply chain, they are indirectly concerned as well and have to provide information about their activities, too. Thereby, the challenge arises that firms often can only control and gain information from their direct corporate partners, which prevents a completely transparent supply chain. Schwarzkopf et al. [33] recommend using blockchain technology to comply with the legislative requirements to create a transparent supply chain that delivers extended information from all companies affected in the supply chain that exceed information from direct suppliers.

4.3 Product-related Blockchain-based Use Cases

One widespread proposed application area for blockchain-based systems is to track and trace food and deliver provenance data for customers such as in [34], [37–40]. For instance, in a study by IBM [34], two blockchain-based food traceability systems for pork and mangos were developed. Such systems should hinder food scandals and create trust in food products. Further, a retailer, in this case Walmart, can deliver additional services to their customers that create a competitive advantage.

Tian [37] conceptualizes a blockchain-based framework for an agri-food supply chain traceability system where Radio Frequency Identification (RFID) realizes the track and traceability system. This blockchain approach has several advantages compared to a centralized system combined with RFID. Firstly, due to the blockchain's decentralization every authorized actor in the supply chain can read and save data which creates higher flexibility to include new actors. Secondly, the decentralization of the blockchain leads technically to a non-existence of institutions, e.g. government departments or companies, and as a result no frauds or corruptions can happen. Thirdly, tracked and traced foods that have a unique ID in the RFID system can be used to protect against imitations. A similar approach like in [37] suggests Biswas et al. [39]

for a wine supply chain traceability system. Mentionable is that they identified mostly the same reasons for the implementation of a blockchain-based system to prevent imitations, food chemicals and in general bad handling of the products.

Besides food, the automotive industry is another suitable application area for blockchain-based systems. Notheisen et al. [42] developed a system to trade cars on the blockchain, more precisely they choose the domain of the Danish Motor Register. The solution tries to stop the adverse selection effect for used cars by recording the history of the vehicles. Moreover, it aims to replace today's institutional centralized registers with a decentralized autonomous system. A slightly different use case present Hua et al. [35], who implemented a blockchain-based system to track the status of batteries in electric vehicles to enable a battery swapping mode that allows a fast battery "refueling". In this concept trust regarding the battery's status and history is essential. According to Hua et al. [35] blockchain is appropriate to create the necessary trust between the electronic vehicle driver and the battery swapping provider. In addition, it allows correct billing due to the record of both batteries.

Another field of application, which has an interest in blockchain technology is the aviation industry. The supply chain in this industry is complex due to globally widespread assembly hubs. Further, systematic capturing of the condition and lifetime expectations from the observed parts is beneficial when scheduling airplane maintenance or overhaul. Besides a transparent supply chain, blockchain leads to a proper monitoring and prevents black market [31]. Also, Toyoda et al. [32] encountered imitations with their conceptualized and implemented "Product Ownership Management System (POMS)" that is based on blockchain and specially designed for processes after the first sale e.g. reselling. Marfia and Esposti [36] suggest using blockchain technology and sensors to elicit the quality of products for customers. The system should replace vulnerable certification systems and increase or stabilize the customer's willingness-to-pay. As a result, reshoring activities of businesses should grow because the customer can gain transparency and evidence regarding excellent quality.

4.4 Service and Sharing Economy-related Blockchain-based Use Cases

Although the aforementioned blockchain applications have a product-centric focus, they allow offering special services that are related to the product. Seebacher and Schürtz [3] carried out a systematic literature review to investigate the applicability of blockchain for service systems and found a high impact of blockchain to service systems. In their point of view, the blockchain characteristics such as trusted environment and decentralization are especially suitable to improve service systems. Hans et al. [43] focus in their study on insurances: One application area is to enhance the contract management for catastrophe swaps (CAT swaps) in the domain of natural catastrophe insurances. Smart contracts and the underlying blockchain technology can represent CAT swaps and trigger specific processes if the negotiated conditions are fulfilled. The second use case is within day-based insurances that can be requested for special circumstances e.g. time-based flight insurance against delayed or canceled flights. A sharing economy relies on information exchanges, Pazaitis et al. [28]

developed the framework named Backfeed that uses blockchain to realize decentralized cooperation. Thereby the framework addresses a more efficient approach for value exchanges in the sharing economy. Hawlitschek et al. [29] explore the effects of blockchain in the sharing economy, too. They discovered the challenges that arise with their implementation. Despite blockchain technology should deliver trust-free systems – the concept is not fully applicable within the sharing economy. A shift from trust in institutions into trust in algorithms was observed. Additionally, in their point of view trust build upon technology is not sufficient and therefore propose a supplementary feedback system.

4.5 Supply Chain-related Blockchain-based Use Cases

In a Delphi study, White [44] figured out that supply chain management systems and independent certification of product quality are possible blockchain-based applications in the future. Overall, the use case of supply chain management systems based on blockchain is evaluated with four of the most probable application scenarios with the highest impact in their Delphi study. Korpela et al. [45] investigate the application of blockchain into supply chains. In focus groups with business experts they elicit requirements and evaluate the applicability of blockchain within digital supply chains. They discovered that several of today's unsolved business requirements e.g. to accelerate digital supply chains could be met by blockchain technology. In a consortium research project, Sternberg and Baruffaldi [46] conceptualized and developed the ReLog concept that saves supply chain information into the Blockchain and thereby collects data along the value chain. Kim and Laskowski [38] try to enhance the specification of blockchain-based applications with ontologies. A formal ontology is used for the design of the blockchain and smart contracts. They developed a proof-of-concept in the domain of supply chains to realize traceability for products.

5 Blockchain-based Use Cases for Sustainable Product-Service Systems

We address the second research question with a product-life cycle model in accordance to Blinn et al. [9] to illustrate the impact of the blockchain technology. The presented blockchain-based product-life cycle model (cf. figure 4) is based on the identified blockchain-based use cases (cf. section 4) and explains the influence of blockchain on product consumption and recycling activities. Hence the model explains what kind of blockchain use cases cause sustainable effects on PSS. The product-life cycle is divided on the horizontal axis into RFID/IoT supply chain and post supply chain according to Toyoda et al. [32]. Thus the switch occurs when the product is first sold. On the vertical axis, the lower quadrants illustrate efforts to implement blockchain with track and tracing systems, e.g. RFID or IoT and to realize PSS. The upper quadrants constitute the perceived value of the entire solution, i.e. the blockchain system in combination with PSS. Due to the application of blockchain, inherent effects will be achieved that have an impact on PSS usage. In the process, the perceived PSS value will be stretched and increased by the application of blockchain, which results in an improved resource efficiency that supports the circular economy. The following concepts influence the graph of the perceived value:

A) *Provenance and Product History:* The perceived value of products and services will be increased because the customer's retrieve transparent information about the product assuming that the declared product quality will be reflected in the data. The provision of information reduces information asymmetry and returns trust in products that got lost in the past e.g. in the food industry [34], [39]. Furthermore, it prevents black market [31] and thereby reduces potential customer's distrust again. Overall to provide information about the origin of the product results in a higher willingness-to-pay for high-quality products [36].

B) *Life Time Expiration:* The lifetime of PSS will be stretched. Within reselling or sharing activities, we assume that no decrease of the perceived PSS value will take place. This can be explained by transparency and information exchange that reduces adverse selection [42] when products are resold or to track the estimated consumption when products are shared [35]. Further, the provision of information for maintenance [31] which is also imaginable for self-service and smart contracts that extends guarantees [43] can stretch the product lifetime.

C) *Waste Management:* The recycling phase occurred as a result of decreased perceived value during the lifetime due to consumption, aging or abrasion and the perceived PSS value is almost equal to the material value. Especially for complex PSS blockchain can support the waste management process [8] and ensure a higher probability to reuse product components [7]. Furthermore, product stewardship will increase because of information transparency, i.e. collecting historical data, that results in the design of more recyclable products.

D) *Track and Tracing Supply Chain:* The overall track and tracing of products in the supply chain will decrease frauds [31] and bad behavior [39] in PSS. In contrast, we expect a higher probability of good behavior and extra efforts to ensure high-quality outcomes.

E) *Data Analysis for Product Design:* The collected product-life cycle data can be used during the product design phase to receive new insights and to create more durable and sustainable PSS [31], [41].

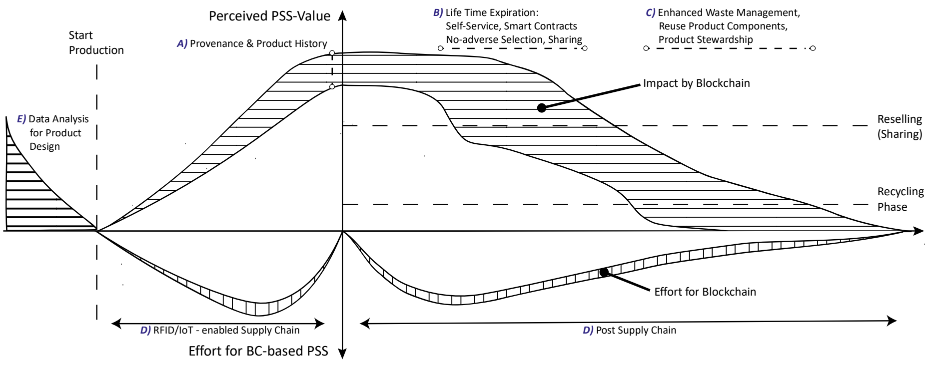


Figure 4. Blockchain-based Product-Life Cycle Model in consideration of Perceived Value

6 Limitations, Further Research, and Conclusion

6.1 Limitations and Further Research

We are aware of the limitations in our research. Therefore, we mention further research topics to overcome the existing limitations. The paper contributes an initial step and combines the three subjects blockchain, PSS and circular economy. Thereby, one limitation is that the findings are mainly based on academic contributions and theoretical assumptions. These expectations have to be validated with further research that is focusing on practical evidence and implementation. Further, the identified blockchain use cases resulting from different domains. Therefore, it is questionable if all use cases can affect each PSS in the same way, i.e. the limitation is an abstract perspective. Furthermore, the developed blockchain-based product-life cycle model is based on scientific literature but was conceptualized in a subjective view by the authors. However, the model serves to understand the different impacts of blockchain on the product-life cycle and how technology enhances resource utilization related to the circular economy. Future long-term studies (5 – 10 years) have to investigate each mentioned impact with practical blockchain implementations. During the studies, additional requirements from practitioners ought to be included and the existing list of blockchain-based use cases should be extended and prioritized. Additionally, investigations have to be made regarding the feasibility – not every product category is suitable to record the product-life cycle and estimated costs have to be relatively low. For instance, Toyoda et al. [32] assumes 1$ for six transfer actions with their blockchain solution and suggest to apply products above 100$. Finally, deeper research how blockchain, IoT, PSS, and circular economy relate together should be conducted.

6.2 Conclusion

The circular economy faces different implementation barriers [2]. We address these obstacles with the combination of blockchain and PSS. Thereby blockchain-based advances such as trust, information exchange, and immutability enable new potentials to solve current issues in the circular economy domain. PSS serves as the application area due to already proven positive effects on sustainability [6]. Based on a conducted literature review we identified 14 unique blockchain-based use cases that have a sustainable effect on PSS that are related to the circular economy. Hence, the paper presents for the first time to the best of our knowledge a holistic view on blockchain, PSS and circular economy.

The evolved blockchain-based product-life cycle model in consideration of perceived value illustrates the positive impact of blockchain on resource utilization. In addition, the model explains how a more effective information provision by blockchain cause a higher perceived value of products and services during the whole product life for society, which results in improved resource utilization and a realization of the circular economy. Thus, the paper contributes new insights how blockchain can be used to gain ecological and economic benefits through a new digital technology.

7 Acknowledgements

This research is partly funded by the European Regional Development Fund and the State of Lower Saxony (NBank) in the scope of the research project SmartHybrid – Service Engineering (ID: 6-85003236). We would like to thank them for their support.

References

1. Geissdoerfer, M., Savaget, P., Bocken, N.M.P., Hultink, E.J.: The Circular Economy – A new sustainability paradigm?. J. Clean. Prod. 143, 757–768 (2017)
2. Darla, G., Galvão, A., Nadae, J. De, Honorato, D., Stief, P., Dantan, J., Etienne, A., Siadat, A.: Circular Economy : Overview of Barriers. Procedia CIRP 73, 79–85 (2018)
3. Seebacher, S., Schüritz, R.: Blockchain Technology as an Enabler of Service Systems: A Structured Literature Review. In: Za, S., Druagoicea, M., and Cavallari, M. (eds.) Exploring Services Science. pp. 12–23. Springer International Publishing, Cham (2017)
4. Gartner: Top Trends in the Gartner Hype Cycle for Emerging Technologies, 2017 - Smarter With Gartner, https://www.gartner.com/smarterwithgartner/top-trends-in-the-gartner-hype-cycle-for-emerging-technologies-2017/ (Accessed: 20.11.2018)
5. Bitcoin Energy Consumption Index – Digiconomist, https://digiconomist.net/bitcoin-energy-consumption (Accessed: 20.11.2018)
6. Hüer, L., Hagen, S., Thomas, O., Pfisterer, H.-J.: Impacts of Product-Service Systems on Sustainability – A structured Literature Review. Procedia CIRP 73, 228–234 (2018)
7. Saberi, S., Kouhizadeh, M., Sarkis, J.: Blockchain technology: A panacea or pariah for resources conservation and recycling?. Resour. Conserv. Recycl. 130, 80–81 (2018)
8. Ongena, G., Utrecht, H., Smit, K.: Blockchain-based Smart Contracts in Waste Management : A Silver Bullet ? In: 31st Bled eConference: Digital Transformation: Meeting the Challenges. pp. 345–355 (2018)
9. Blinn, N., Nüttgens, M., Schlicker, M., Thomas, O., Walter, P.: Lebenszyklusmodelle hybrider Wertschöpfung: Modellimplikationen und Fallstudie. In: Thomas, O., Loos, P., and Nüttgens, M. (eds.) Hybride Wertschöpfung. pp. 130–143. Springer, Berlin Heidelberg, (2010)
10. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., Brocke, J. Von, Reimer, K.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. 17th Eur. Conf. Inf. Syst. Paper 161, 2206–2217 (2009)
11. Dybå, T., Dingsøyr, T.: Empirical studies of agile software development: A systematic review. Inf. Softw. Technol. 50, 833–859 (2008)
12. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, 13–23 (2002)
13. Nakamoto, S.: Bitcoin : A Peer-to-Peer Electronic Cash System. 1–9 (2008)
14. Nofer, M., Gomber, P., Hinz, O., Schiereck, D.: Blockchain. Bus Inf Syst Eng. 59, 183–187 (2017)
15. Li, Y., Marrie-Bienvenue, T., Perron-Brault, A., Wang, X., Paré, G.: Blockchain Technology in Business Organizations : A Scoping Review. In: Proceedings of the 51st Hawaii International Conference on System Sciences. pp. 4474–4483 (2018)
16. Konstantinidis, I., Siaminos, G., Timplalexis, C., Zervas, P., Peristeras, V., Decker, S.: Blockchain for Business Applications: A Systematic Literature Review. In: Abramowicz,

W. and Paschke, A. (eds.) Business Information Systems. pp. 384–399. Springer International Publishing, Cham (2018)
17. Witt, J., Richter, S.: Ein problemzentrierter Blick auf Blockchain- Anwendungsfälle. In: MKWI 2018 Lüneburg. pp. 1247–1258 (2018)
18. Risius, M., Spohrer, K.: A Blockchain Research Framework: What We (don't) Know, Where We Go from Here, and How We Will Get There. Bus. Inf. Syst. Eng. 59, 385–409 (2017)
19. Mont, O.: Clarifying the concept of product – service system. J. Clean. Prod. 10, 237–245 (2002)
20. Thomas, O., Walter, P., Loos, P.: Product-Service Systems: Konstruktion und Anwendung einer Entwicklungsmethodik. Wirtschaftsinformatik 50, 208–219 (2008)
21. Aurich, J.C., Fuchs, C., Wagenknecht, C.: Life cycle oriented design of technical Product-Service Systems. J. Clean. Prod. 14, 1480–1494 (2006)
22. Cavalieri, S., Pezzotta, G.: Product–Service Systems Engineering: State of the art and research challenges. Comput. Ind. 63, 278–288 (2012)
23. Morelli, N.: Product-service systems, a perspective shift for designers: A case study: the design of a telecentre. Des. Stud. 24, 73–99 (2003)
24. Goedkoop, M.J., Van Halen, C.J.G., Te Riele, H.R.M., Rommens, P.J.M.: Product Service systems , Ecological and Economic Basics. (1999)
25. Özcan, D., Kammler, F., Thomas, O.: Integrationsansatz zum konzeptionellen Design von Product-Service Systems. In: Thomas, O. and Nüttgens, M. (eds.) Dienstleistungsmodellierung 2014. pp. 54–75. Springer Fachmedien, Wiesbaden (2014)
26. Tukker, A.: Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet. Bus. Strateg. Environ. 13, 246–260 (2004)
27. Tukker, A.: Product services for a resource-efficient and circular economy – A review. J. Clean. Prod. 97, 76–91 (2015)
28. Pazaitis, A., Filippi, P. De, Kostakis, V.: Blockchain and value systems in the sharing economy: The illustrative case of Backfeed. Technol. Forecast. Soc. Change. 125, 105–115 (2017)
29. Hawlitschek, F., Notheisen, B., Teubner, T.: The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy, 29, 1–41 (2018)
30. Askoxylakis, I., Alexandris, G., Demetriou, G.: Blockchain-enabled Intelligent Asset Exchange for a Circular Economy. ERCIM News July (2017)
31. Madhwal, Y., Panfilov, P.B.: Industrial Case : Blockchain on Aircraft ' s Parts Supply Chain Management. Am. Conf. Inf. Syst. 2017 Work. Smart Manuf. Proc., pp. 1–6 (2017).
32. Toyoda, K., Mathiopoulos, P.T., Sasase, I., Ohtsuki, T.: A Novel Blockchain-Based Product Ownership Management System (POMS) for Anti-Counterfeits in the Post Supply Chain. IEEE Access. 5, 17465–17477 (2017)
33. Schwarzkopf, J., Adam, K., Wittenberg, S.: Vertrauen in nachhaltigkeitsorientierte Audits und in Transparenz von Lieferketten – Schafft die Blockchain-Technologie einen Mehrwert? In: Khare, A., Kessler, D., and Wirsam, J. (eds.) Marktorientiertes Produkt- und Produktionsmanagement in digitalen Umwelten: Festgabe für Klaus Bellmann zum 75. Geburtstag. pp. 171–180. Springer Fachmedien Wiesbaden, Wiesbaden (2018)
34. Kamath, R.: Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM. J. Br. Blockchain Assoc. 1, 1–12 (2018)
35. Hua, S., Zhou, E., Pi, B., Sun, J., Nomura, Y., Kurihara, H.: Apply blockchain technology to electric vehicle battery refueling. Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 4494–4502 (2018)

36. Marfia, G., Esposti, P.D.: Blockchain and Sensor-Based Reputation Enforcement for the Support of the Reshoring of Business Activities. In: Vecchi, A. (ed.) Reshoring of Manufacturing: Drivers, Opportunities, and Challenges. pp. 125–139. Springer International Publishing, Cham (2017)
37. Tian, F.: An Agri-food Supply Chain Traceability System for China Based on RFID & Blockchain Technology. (2016).
38. Kim, H., Laskowski, M.: Agriculture on the Blockchain: Sustainable Solutions for Food, Farmers, and Financing. Blockchain Research Institute, York University (2017)
39. Biswas, K., Muthukkumarasamy, V., Tan, W.L.: Blockchain Based Wine Supply Chain Traceability System. Future Technologies Conference (FTC), pp. 1–7 (2017)
40. Bettín-Díaz, R., Rojas, A.E., Mejía-Moncayo, C.: Methodological Approach to the Definition of a Blockchain System for the Food Industry Supply Chain Traceability. In: Gervasi, O., Murgante, B., Misra, S., Stankova, E., Torre, C.M., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O., Tarantino, E., and Ryu, Y. (eds.) Computational Science and Its Applications ICCSA 2018. pp. 19–33. Springer International Publishing, Cham (2018)
41. Düring, T., Fisbeck, H.: Einsatz der Blockchain-Technologie für eine transparente Wertschöpfungskette. In: Hildebrandt, A. and Landhäußer, W. (eds.) CSR und Digitalisierung: Der digitale Wandel als Chance und Herausforderung für Wirtschaft und Gesellschaft. pp. 449–464. Springer, Berlin, Heidelberg (2017)
42. Notheisen, B., Cholewa, J.B., Shanmugam, A.P.: Trading Real-World Assets on Blockchain: An Application of Trust-Free Transaction Systems in the Market for Lemons. Bus. Inf. Syst. Eng. 59, 425–440 (2017)
43. Hans, R., Zuber, H., Rizk, A., Steinmetz, R.: Blockchain and Smart Contracts: Disruptive Technologies for the Insurance Market. In: Twenty-third Americas Conference on Information Systems. pp. 1–10. , Boston (2017)
44. White, G.R.T.: Future applications of blockchain in business and management : A Delphi study. Strategic Change 26, 439–451 (2017)
45. Korpela, K., Hallikas, J., Dahlberg, T.: Digital Supply Chain Transformation toward Blockchain Integration. In: Proceedings of the 50th Hawaii International Conference on System Sciences. pp. 4182–4191 (2017)
46. Sternberg, H., Baruffaldi, G.: Chains in Chains – Logic and Challenges of Blockchains in Supply Chains. In: Proceedings of the 51st Hawaii International Conference on System Sciences. pp. 3936–3943 (2018)
47. European Parliament: EU-Richtlinie 2014/95/EU. Off. J. Eur. Union. 2014, 1–9 (2014)

Digitale Rückverfolgbarkeit von Lebensmitteln: Eine verbraucherinformatische Studie

Stephanie Vonholdt[1], Gunnar Stevens[2], Karoline Kleih[1], und Alexander Boden[3]

[1] Hochschule Bonn-Rhein-Sieg, Wirtschaftswissenschaften, Sankt Augustin, Deutschland
Stephanie.Vonholdt@h-brs.de, karoline.kleih@smail.wis.h-brs.de
[2] Universität Siegen, IT-Sicherheit und Verbraucherinformatik, Siegen, Deutschland
Gunnar.Stevens@uni-siegen.de
[3] Fraunhofer Institute for Applied Information Technology FIT, Department of User-Centered Computing, Sankt Augustin, Deutschland
alexander.boden@fit.fraunhofer.de

Abstract. Die Globalisierung führt zu immer komplexeren, für die Einzelnen kaum nachvollziehbaren Wertschöpfungsketten in der Lebensmittelindustrie. Zugleich eröffnet die Digitalisierung neue Möglichkeiten, Informationen entlang der Kette zu sammeln, und so mehr Transparenz und Vertrauen für den Verbraucher beziehungsweise die Verbraucherin zu schaffen. Jedoch finden Verbraucherinformations-Apps wie *fTRACE* bisher nur eine geringe Verbreitung. Daher haben wir in einer qualitativen Studie mit 16 Teilnehmer/-innen Bedürfnisse und Nutzungshürden von Verbraucher/-innen im Zusammenhang mit Verbraucherinformations-Apps analysiert. Es zeigt sich, dass das Vertrauen in die Informationen, sowie der einfache Zugang dazu für Verbraucher/-innen zentral sind. Durch die gut sichtbare Bereitstellung der Informationen am Point-of-Sale, sowie der automatisierten Informationsversorgung z. B. mittels digitaler Kassenzettel in Kombination mit weiteren Verbraucher-Services kann die Bekanntheit und Akzeptanz von Rückverfolgbarkeitssystemen weiter gesteigert werden.

Keywords: Traceability, Lebensmittelindustrie, Supply Chain Management, Verbraucherinformatik.

1 Einleitung

Durch die komplexer werdenden Wertschöpfungsnetzwerke einer globalisierten Nahrungsproduktion und -distribution, das Aufkommen genetisch veränderter Organismen (GVO), sowie wiederkehrende Lebensmittelskandale (BSE, „Gammelfleisch“, Pferdefleisch in als Rindfleischprodukten deklarierten Lebensmitteln, etc.) steigt die Nachfrage nach einer besser nachprüfbaren Qualität und Sicherheit von Lebensmitteln. Insbesondere bei tierischen Produkten ist das Wissen um eine tiergerechte Haltung, saubere Verarbeitung und Frische zu einem wichtigen Kaufkriterium für Verbraucher/-innen geworden. Um dem Wunsch der Verbraucher/-innen nach einer gleichbleibenden Versorgung mit hochwertigen, sicheren und nährstoffreichen Lebensmitteln gerecht zu

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

werden, sowie für den Wiederaufbau von Vertrauen der Öffentlichkeit in die Nahrungsmittelkette, bedarf es der Gestaltung und Umsetzung vollständig rückwärts und vorwärts verfolgbarer Lieferketten vom Erzeuger bis zum/zur Verbraucher/-in. In den letzten Jahren wurden zunehmend digitalisierte Rückverfolgbarkeitssysteme in der Lieferkette der Lebensmittel- und Agrarwirtschaft etabliert [1]. Während diese im Business-to-Business Bereich relativ gut erforscht wurden [2, 3], gibt es bisher kaum Studien zur „letzten Meile" hin zum/zur Endverbraucher/-in. Der vorliegende Beitrag analysiert deshalb die Bedarfe, die Nutzung und Akzeptanz von Rückverfolgbarkeitssystemen aus Verbrauchersicht. Hierzu stellen wir in unserem Beitrag den aktuellen Forschungsstand und die Ergebnisse einer qualitativen Studie vor, die Bedarfe und Nutzungshürden von Verbraucher/-innen auf Basis des Systems *fTRACE* analysiert, um daraus Anforderungen zur Verbesserung der Rückverfolgbarkeit aus Verbrauchersicht zu erarbeiten.

2 Rückverfolgbarkeit von Lebensmitteln

Die von Endverbraucher/-innen konsumierten Lebensmittel sind das Ergebnis eines langen Produktionsprozesses, der über mehrere Wertschöpfungsstufen verläuft [2]. Rückverfolgbarkeit meint dabei die Fähigkeit, Ereignisse in allen Stufen einer Prozesskette chronologisch zu erfassen und zu überprüfen [4]. Nach der EU Verordnung 178/2002 umfasst dies bei Lebensmitteln alle Produktions-, Verarbeitungs- und Vertriebsstufen [5].

Ziele der Rückverfolgbarkeit sind u. a. die Optimierung von Produktions- und Logistikprozessen, die Verbesserung der Lebensmittelsicherheit, die Ursachenforschung z. B. bei Verkeimung, sowie die Ermöglichung eines Rückrufs von bestimmten Produktchargen [1, 3, 6–8]. In einer Literaturübersicht haben Karlsen et al. [9] weitere Treiber zur Umsetzung von Rückverfolgbarkeitssystemen identifiziert, wie rechtliche Anforderungen, Nachhaltigkeitsanforderungen, Zertifizierungsanforderungen, Erlangung eines Wettbewerbsvorteils oder Schutz vor biologischen Anschlägen. Verschiedene Studien belegen, dass eine Reihe von Faktoren bei der Umsetzung eine Rolle spielen. Eine hohe Herausforderung besteht dabei in der Komplexität von Wertschöpfungsnetzwerken [4]. Dabei sind Kosten und Nutzen von Rückverfolgbarkeitssystemen in der Wertschöpfungsketten nicht immer gleich verteilt [9]. Die Motivation zur Einführung bzw. Teilnahme an einem Rückverfolgbarkeitssystem ist daher eng verbunden mit der Identifizierung individueller Kosten- und Nutzenfaktoren bei den einzelnen Akteuren innerhalb der Wertschöpfungskette [10]. Rückverfolgbarkeit ist deshalb vor allem eine verteilte Managementaufgabe, bei der die Aktivitäten verschiedener Akteure koordiniert werden müssen [2, 8, 9].

Zentral für die Implementierung einer Rückverfolgbarkeit ist eine effektive und effiziente Informationslogistik, da vollständige Informationen über Prozessketten notwendig sind, um einerseits die Einhaltung der Spezifikationen zu überprüfen, und andererseits die Ursachen von Ausfällen und unerwünschten Ereignissen zu identifizieren [3]. Verschiedene Autoren sehen dabei die Digitalisierung als Kernelement für die signifikante Verbesserung der Rückverfolgbarkeit von Produkten [3, 11, 12]. Im Wesentlichen basieren Rückverfolgbarkeitssysteme technisch auf vier Säulen [3]: 1) Merkmale zur eindeutigen Produktidentifikation, 2) Daten zur Rückverfolgbarkeit entlang von Wertschöpfungsketten, 3) Unterstützung von Produktrouting und 4) Werkzeuge und Standards zur Rückverfolgung.

In der Industrie haben vor allem die Standards der GS1 (Global Standards One) eine weite Verbreitung gefunden [13–15]. Zur Synchronisation von Stammdaten über die einzelnen Stufen der Wertschöpfungskette gibt es z. B. das Global Data Synchronisation Network (GDSN), welches den automatischen elektronischen Austausch von standardisierten Produktinformationen zwischen Handelspartnern ermöglicht. Mittels Electronic Data Interchange (EDI) können elektronische Transaktionsdaten zwischen Handelspartnern ausgetauscht werden. Hierfür werden die Informationen aus dem Warenwirtschaftssystem vor der Übertragung in ein Standardformat (GS1 XML, EANCOM) konvertiert [13]. Um die Rückverfolgbarkeit von Produkten und Transparenz über die gesamte Lieferkette hinweg gewährleisten zu können, sind dabei insbesondere Ereignisdaten hochrelevant. Diese fallen bei der Auslösung bestimmter Prozesse an, wie beispielsweise der Auslieferung durch einen Logistikdienstleister. Durch die Aufzeichnung von Ereignissen wird die detaillierte Verfolgung von Prozessen in Echtzeit ermöglicht. Die erfassten Ereignisdaten können dann mithilfe des Schnittstellenstandards EPCIS (Electronic Product Code Information Service) verarbeitet und weitergeleitet werden [13].

2.1 Verbraucherinformationen

Die einfache Verfügbarkeit vertrauenswürdiger Verbraucherinformationen stellt eine wesentliche Säule des Verbraucherschutzes dar [16]. Der besondere Schutzbedarf findet seinen Niederschlag z. B. in diversen Verbraucherschutzgesetzen und -verordnungen, wie z. B. der EU-Verordnung Nr. 1169/2011 zur Kennzeichnung von Lebensmitteln. Die Informationen stammen entweder freiwillig oder gesetzlich verpflichtend vom Hersteller, sowie von Verbraucherinstitutionen und Verbrauchermedien.

Die Informationen lassen sich in Bezug auf ihre Eigenschaften wie folgt unterscheiden [16, 17]: 1. *Suchinformationen*, die der/die Käufer/-in vor dem Erwerb eines Produkts durch die eigene Wahrnehmung identifizieren und beurteilen kann; 2. *Erfahrungsinformationen,* die erst durch den Gebrauch eines Produkts nach dem Kauf beurteilt werden können; 3. *Vertrauensinformationen*, die nur langfristig nach dem Kauf eines Produktes beurteilt werden können. Der Käufer oder die Käuferin muss hier darauf vertrauen, dass ein Produkt tatsächlich die gewünschten Eigenschaften erfüllt. Bei der Rückverfolgbarkeit handelt es sich meist um Vertrauensinformationen, da die gesammelten Informationen entlang der Produktions-, Verarbeitungs- und Vertriebsstufen vom Verbraucher beziehungsweise der Verbraucherin selbst nicht überprüft werden können. Hierdurch steigen die Anreize für opportunistisches Verhalten auf Anbieterseite, da bei zunehmender Informationsasymmetrie die Wahrscheinlichkeit von unbeobachteten Qualitätsverschlechterungen steigt.

Verschiedene Studien haben die positive Wirkung von Verbraucherinformationen auf das Verbraucherverhalten gezeigt, insbesondere beim Thema Food Labels [18]. Die Nutzung und das Verständnis von Food Labels hängt jedoch von einer Reihe soziodemographischer Faktoren wie Alter, Einkommen, Bildung und Lebensstil ab [19]. Ferner zeigen Studien dass Verbraucherinformationen zum Teil falsch interpretiert werden, zu komplex sind bzw. darunter leiden, dass sie für Nicht-Experten konzipiert wurden [18, 20]. Hinzu kommt, dass Lebensmittel ein vielschichtiges Qualitätsbündel bestehend aus einer ganzen Reihe von Teileigenschaften darstellen.

Aus Verbrauchersicht sind diese einzelnen Eigenschaften oft nur schwierig oder gar nicht überprüfbar. Daher besteht die Gefahr, dass Verbraucher/-innen durch die Aufmachung der Produkte oder durch Werbung in Bezug auf die tatsächlichen Eigenschaften getäuscht werden [21]. Die Zunahme an Informationen führt deshalb nicht automatisch dazu, dass Verbraucher/-innen besser informiert sind [22]. Bei der Gestaltung von Verbraucherinformationen gilt es deshalb, das Ziel objektiver und vertrauenswürdiger Informationen mit dem Ziel leicht verständlicher, ästhetisch-emotional ansprechender und handlungsorientierter Informationen aus Verbrauchersicht in Einklang zu bringen [23].

2.2 Verbraucher-Apps zur Rückverfolgbarkeit

Der wichtigste Informationskanal für Verbraucher/-innen ist nach wie vor die Kennzeichnung auf Lebensmittelverpackungen, während z. B. die von Behörden bereitgestellten Informationen sowie Beratungsmöglichkeiten im Handel kaum genutzt werden [16]. Darüber hinaus gewannen digitale Verbraucherinformationen wie z. B. Preis- und Vergleichsportale in den letzten Jahren an Bedeutung [24].

Die steigende Verbreitung von Smartphones erlaubt dabei die Entwicklung mobil nutzbarer Verbraucherinformationssysteme (VIS) zur Abfrage von Lebensmittelinformationen [25]. So haben etwa González-Miranda et al. [26] einen VIS-Prototypen vorgestellt, der Kundenprofile nutzt, um passgenauere Verbraucherinformationen zur Verfügung zu stellen. Das Konzept my2cents der ETH Zürich stellt eine mobile VIS-Anwendung dar, bei der Verbraucher/-innen den Barcode einscannen können, um Kommentare und Meinungen zu Produkten teilen zu können [27]. Stevens et al. [28, 29] stellen ein Konzept eines digitalen Kassenzettels vor, mittels dessen Verbraucher/-innen kontextspezifische, sowie aggregierte Informationen zum Einkauf bekommen können. Ozarslan und Eren stellen ein kontextbasiertes VIS-Konzept vor, das Verbraucher/-innen entlang des Entscheidungsprozesses vor, während und nach dem Kauf unterstützen soll [30]. Diese Ansätze unterstützen bisher jedoch noch nicht die Rückverfolgbarkeit von Lebensmitteln.

Daneben gibt es eine Reihe kommerzieller Lösungen, wie z. B. die App barcoo oder Codecheck, bei der der Nutzer oder die Nutzerin den Barcode eines Produkts einscannt, um Informationen zu Inhaltsstoffen, Lebensmittel-Ampeln, Preisvergleiche, sowie Testberichte und Nutzerrezensionen angezeigt zu bekommen. Einen etwas anderen Ansatz hat die Web-Plattform WikiProducts, die ein offenes, nicht moderiertes Forum bietet, in welchem Informationen zu Produkten gesammelt und strukturiert werden. Neben Verbraucher/-innen, Herstellern und Händlern zählen auch Universitäten und nichtstaatliche Organisationen (NGOs) zu den beteiligten Akteuren. Die genannten Lösungen unterstützen bisher jedoch ebenfalls nicht die digitale Rückverfolgung von Produkten, insbesondere von Lebensmitteln.

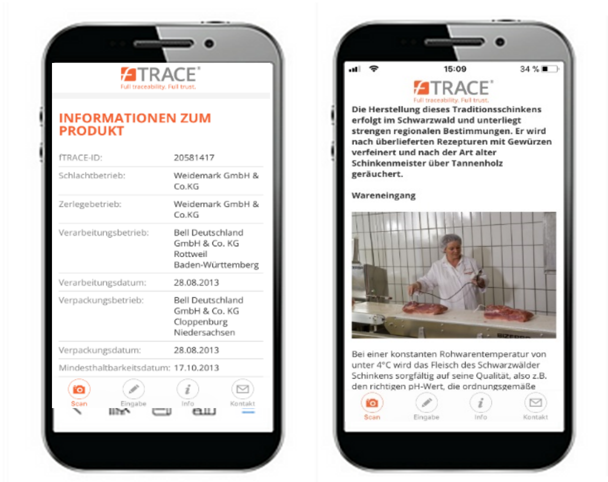


Abbildung 1: Screenshot der Rückverfolgbarkeits-App *fTRACE*

Einen Ansatz für Rückverfolgung bietet derzeit vor allem die cloudbasierte Traceability-Plattform *fTRACE*, die technisch auf den oben skizzierten Standards der GS1 aufsetzt [31]. Die Verbraucher/-innen können mit der *fTRACE* App den auf der Verpackung angebrachten QR-Code (in den DataMatrix, d. h. GTIN, Chargennummer plus Zusatzangaben kodiert ist) einscannen, um Informationen zum Produkt abzurufen (vgl. Abbildung 1). Die Daten zur Herkunft, Verarbeitung und Produktqualität werden dabei aus erster Hand von Herstellern und Händlern bereitgestellt, welche auf diesem Weg auch häufig freiwillige Informationen zur Qualitätsüberwachung oder Rezepttipps zu ihren Produkten anbieten.

3 Methodik

Um die Bedarfe von Verbraucher/-innen und die Kontexte zur Nutzung von Werkzeugen zur digitalen Rückverfolgbarkeit von Lebensmitteln genauer zu verstehen, haben wir 16 problemzentrierte Interviews durchgeführt [32]. Teilnehmer/-innen wurden nach dem Schneeballsystem ausgewählt (Alter: 17 - 59, Ø: 34,4; männlich: 9, weiblich: 7; 9 Berufstätige, 6 Studierende, 1 Schüler). Die durchschnittliche Länge der Interviews betrug 23 Minuten. Zwölf Interviews wurden zur Auswertung vollständig transkribiert, in vier Fällen wurde das Gespräch aus technischen oder persönlichen Gründen der Teilnehmer/-innen nicht aufgezeichnet und stattdessen ein ausführliches Feldprotokoll angefertigt.

Die Struktur der Interviews war zweigeteilt: Im ersten Teil wurden die Teilnehmer/-innen anhand eines offenen Gesprächsleitfadens zu ihrem Einkaufsverhalten, ihrem Umgang mit Verbraucherinformationen und zum Konzept der digitalen Rückverfolgbarkeit von Lebensmitteln befragt. Die Interviewsituation hatte dabei einen offenen Charakter, um so den Teilnehmer/-innen Raum für das Äußern eigener Begründungen und Meinungen zu geben. Teilweise wurde das Gespräch jedoch wieder auf die zugrundeliegende Problemstellung zurückgeführt, um eine Strukturierung zu erzielen.

Im zweiten Teil der Studie wurde die oben beschriebene führende Rückverfolgbarkeits-App *fTRACE* im Sinne eines Design Probe genutzt, um durch die Nutzung der App konkrete Gestaltungsprobleme zu identifizieren, aber insbesondere die Reflektion über Bedarfe und potentielle Nutzungskontexte anzuregen [33]. Hierzu wurde die Verpackung eines typischen Beispielprodukts zur Verfügung gestellt, welche mit der App eingescannt werden konnte. Aufgrund der Homogenität der durch die App abgedeckten Produktpalette und um die Interviews nicht zu sehr in die Länge zu ziehen, wurde auf einen Vergleich mehrerer Produkte verzichtet. Stattdessen wurde die Interaktion mit dem Beispielprodukt tiefergehend untersucht. So sollten die Befragten ihre Eindrücke zu den angezeigten Informationen, der Gestaltung und der Bedienung der App detailliert im Sinne von „Thinking Aloud" schildern. Abschließend erfolgte eine Diskussion in Bezug auf deren praktische Nutzbarkeit bzw. erwünschte Verbesserungen und Ergänzungen.

Die Transkripte und Feldprotokolle wurden kodiert und hieraus übergeordnete Kategorien abgeleitet. Die Analyse fußte auf der Annahme der konstruktivistischen Grounded Theory [34], dass die Interviews und das hierauf aufbauende Kategoriensystem eine gegenseitige Leistung und Interpretation des Interviewers und des Interviewten darstellen. Zum Aufbau des Kategoriensystems wurden die Textabschnitte relevanten Kategorien zugeordnet, die bei der Vorbereitung der Studie identifiziert wurden und auch im Gesprächsleitfaden ihren Niederschlag gefunden haben. Im Rahmen der Analyse wurde das System iterativ erweitert und angepasst, um so den Sichtweisen der Interviewten besser gerecht zu werden.

Bei der Analyse des ersten Teils der Studie wurde zunächst der allgemeine Umgang der befragten Personen mit Verbraucherinformationen (bezüglich Lebensmitteln) betrachtet. Eine zentrale Kategorie war hier generelles Vertrauen in digitale Informationen, da dies für die Studie eine zentrale Rolle spielte. Weitere zentrale Kategorien waren erwartete Effekte (wie etwa weniger Lebensmittelverschwendung) sowie die generelle Einstellung der befragten Personen zur Rückverfolgbarkeit von Lebensmitteln.

Für den zweiten Teil der Studie wurden die Erwartungen der Befragten vor der Nutzung der App reflektiert. Relevante Kategorien waren hier die Bewertung der angezeigten Informationen und der mögliche Nutzungskontext der Anwendung, sowie die Zufriedenheit mit der Darstellung und Bedienung. Die letzte Kategorie umfasste Aussagen dazu, welche zusätzlichen Funktionen und Mehrwertdienste die Teilnehmer/-innen im Zusammenhang mit entsprechenden Systemen für sinnvoll hielten.

4 Ergebnisse

4.1 Allgemeines Interesse, jedoch geringer Bekanntheits- und Nutzungsgrad

Generell fanden zwölf der 16 Teilnehmer/-innen das Konzept der Rückverfolgbarkeit von Lebensmitteln gut und hilfreich, auch wenn zehn davon beim Einkauf nach eigenen

Angaben bisher nicht explizit auf diese Möglichkeit geachtet haben. Auch die vier Befragten, denen die Möglichkeit der Rückverfolgbarkeit tendenziell nicht so wichtig war, sahen durchaus Vorteile, beispielsweise, dass insgesamt ein besseres Bewusstsein für Lebensmittel und deren Herstellung geschaffen werden könnte. Von vier Teilnehmer/-innen wurde sogar der Wunsch geäußert, dass eine digitale lückenlose Rückverfolgbarkeit generell möglich sein sollte und in gewisser Hinsicht ein Recht des/der Verbraucher/-in darstellt:

> *„Ich find' das sollte eigentlich Standard sein, dass das lückenlos der Verbraucher, wenn er das möchte, abrufen kann. Grade wenn er irgendwelche Bedenken hat, dass er da Schritt für Schritt nachverfolgen kann." (I1, 56, Angestellte im öffentlichen Dienst)*

> *„Fleisch ist da doch wahrscheinlich das Paradebeispiel. Da würd' ich's wissen wollen. Da find' ich, hab' ich ,nen Recht drauf, zu erfahren, wie, wo, was mit dem Tier passiert." (I4, 23, Maschinenbau-Studentin)*

Es zeigte sich, dass der Wunsch nach Rückverfolgbarkeit bei tierischen Produkten, vor allem bei Fleisch und anderen leicht verderblichen Lebensmitteln besonders hoch (13-mal) war. Ebenfalls ausgeprägt (6-mal) war der Wunsch nach einer besseren Nachvollziehbarkeit von im Ausland hergestellten oder verarbeiteten Lebensmitteln, um die wahrgenommenen Unsicherheiten bezüglich solcher Produkte zu mindern.

In den Interviews zeigte sich ferner eine große Diskrepanz zwischen Interesse und tatsächlicher Nutzung der digitalen Rückverfolgbarkeit. Insgesamt war das Wissen zu Möglichkeiten digitaler Rückverfolgbarkeit bei den Verbraucher/-innen eher vage. Selbst die zehn Nutzer/-innen, die davon schon einmal gehört hatten, wussten oft nur, dass bei manchen Produkten Informationen durch das Scannen eines QR-Codes angezeigt werden können. Genutzt hatte diese Möglichkeit jedoch bisher keine/-r der Teilnehmer/-innen:

> *„Davor war's mir nicht bekannt. Also ich wusste, dass es sowas in der Art gibt, aber ich wusste auch nicht, dass zum Beispiel das schon bei großen Supermarktketten wirklich vorhanden ist, sondern ich dacht' das wären halt so kleine Versuchsobjekte, wo man das mal ausprobiert hat." (I5, 21, Maschinenbau-Studentin)*

Die in unserer Studie exemplarisch genutzte App *fTRACE* kannte vor der Befragung keine/-r der befragten Verbraucher/-innen. Diese Unkenntnis der existierenden Möglichkeiten führte unter anderem auch dazu, dass Verbraucher/-innen im ersten Teil der Interviews wiederholt Wünsche nach Innovationen äußerten, die bereits von verfügbaren Apps wie *fTRACE* oder *barcoo* angeboten werden:

> *„Da es im Moment praktisch nicht verfügbar ist, wär das natürlich super, wenn's da ein System geben würde, das grade über Scan von einem Code, eben alle Informationen zu dem Produkt zur Verfügung stellt." (I7, 59, Ingenieur)*

Nachdem die Befragten im Interview auf *fTRACE* aufmerksam gemacht wurden, waren 13 der 16 Teilnehmer/-innen positiv überrascht von der Existenz dieses Werkzeugs. Der geringe Nutzungsgrad hängt demnach vor allem mit dem geringen Bekanntheitsgrad solcher Apps zusammen.

„Also (…) ohne die Befragung hätte es wahrscheinlich noch lange gedauert, bis ich auf die App jetzt vielleicht gestoßen wär'. (…) Also ich glaub' das ist noch nicht so in der Bevölkerung oder in der Öffentlichkeit so richtig angekommen, aber ich fänd's toll, wenn das irgendwie mehr beworben wird und (…) wenn man da halt mehr Zugang zu hätte einfach." (I9, 25, VWL-Student)

An dem Zitat zeigt sich auch die wiederholt geäußerte Kritik an der Lebensmittelindustrie und den Verbraucherorganisationen, bisher das Konzept digitaler Rückverfolgung nicht ausreichend gegenüber den Verbraucher/-innen zu kommunizieren. Hierzu trüge auch bei, dass das *fTRACE*-Logo auf Produkten beim Einkauf nicht groß auffällt und man auch nicht durch andere Hinweise darauf gestoßen wird.

„War mir nicht bekannt, nein. Weil ich auch in den Supermärkten vorab keine Informationen über ähnliche oder andere Systeme direkt am POS [Anm. d. Verf.: Point-of-Sale, d. h. am Verkaufsort] finde." (I12, 34, Logistikplaner)

Hier zeigt unsere Studie ein deutliches Potential einer besseren Sichtbarkeit der verfügbaren Informationssysteme auf den Verpackungen bzw. im Handel.

4.2 Informationsbedarfe, -angebote und -darstellung

Die Bedienung, der Informationsumfang und die Darstellung der getesteten Version der *fTRACE* App (siehe Screenshot Abbildung 1) wurde von zwölf der befragten Teilnehmer/-innen in weiten Teilen gelobt, da die von den Teilnehmer/-innen gewünschten wichtigen Informationen überwiegend vorhanden und übersichtlich gestaltet waren. Vier Befragte fanden allerdings, dass z. B. die Herstellungsangaben zu wenig detailliert und daher nur begrenzt informativ seien. Die Bedienung und Navigation wurde von 13 Teilnehmer/-innen als einfach und intuitiv empfunden. Hierzu trug auch die Scan-Funktion bei. Sie erlaubte den Teilnehmer/-innen, ohne hohen Aufwand Informationen zu einem Produkt abzurufen. Jedoch wurde hier bemängelt, dass man teilweise z. B. beim Wechseln des Reiters ein Produkt mehrfach einscannen musste, bevor es funktionierte.
Bei den Informationsbedarfen äußerten die Befragten im ersten Teil des Interviews, dass sie sich für Herkunft, Tierhaltung, Transport und die Verwendung schädlicher Stoffe interessierten. Sie interessierten sich ebenfalls für eine Möglichkeit zur Rückverfolgung, um die Frische der Produkte besser überprüfen zu können, da dies bisher als eher schwierig empfunden wurde.

„(…) bei Geflügelprodukten, weil da das Thema antibiotischer Einsatz in der Zucht, beziehungsweise die Art der Haltung für mich schon 'nen großen Stellenwert hat, ob ich 'nen Produkt kaufe oder nicht. Zumindest hätte ich gerne umfangreiche Informationen darüber." (I12, 34, Logistikplaner)

Die bereitgestellten Informationstexte wurden von 13 der Befragten als gut verständlich und auch die Einteilung als hilfreich bewertet. Von vier Teilnehmer/-innen wurde jedoch angemerkt, dass die Darstellung zur Produktqualität als zu textreich empfunden wird. Hier wurde z. B. von einem Befragten vorgeschlagen, erst einmal nur eine stichpunktartige Übersicht anzuzeigen. Diese sollte Verweise zu weiteren Quellen beinhalten, um sich bei Bedarf genauer informieren zu können. Ferner wurde aussagekräftiges Bild- und Videomaterial z. B. zu Herstellungsbetrieben gewünscht.

Beim Informationsumfang äußerten sich zehn der Teilnehmer/-innen positiv überrascht, wie umfangreich die Informationen zu den Produkten sind. Dabei wurde besonders die Genauigkeit der Daten gelobt, z. B. zu genauen Schlacht- und Zerlegezeiten in chronologischer Anordnung, da dies einen detaillierten Überblick über den zeitlichen Ablauf der einzelnen Produktionsschritte ermöglicht. Ein/-e Teilnehmer/-in nutzte dies zum Beispiel, um den tatsächlichen Zeitraum zwischen Schlachtung und Verpackung bis zum Verkauf einzuschätzen. Positiv hervorgehoben wurde von fünf Teilnehmer/-innen auch, dass in der App Angaben zum EU-Bio-Siegel und Bio-Kontrollen bereitgestellt wurden. Jedoch wurden auch von sechs Teilnehmer/-innen zusätzliche Angaben vermisst, wie z. B. dem Preis des Produkts, Nährwertangaben, Bewertungen anderer Verbraucher/-innen, etc. Gelobt wurden wiederum vermehrt die Angaben zur Herkunft und deren Darstellung in einer geografischen Karte.

> *„Und ich kann's mir in der Karte anschauen, wie ich das grade seh'. Sehr cool, genau sowas stell' ich mir vor, weil dann kann man auch Entfernungen abschätzen (…)“ (I12, 34, Logistikplaner)*

Die Angaben zur Herstellung wurden von sieben Befragten als relevant eingestuft, besonders bei Lebensmitteln, deren Herstellungsprozess bisher nicht bekannt war. Drei davon empfanden die Angaben jedoch als nur begrenzt informativ. So fehlten laut fünf Teilnehmer/-innen auch weiterführende Informationen zu Erzeugungs- bzw. Aufzuchtort, sowie genauere Informationen zu den einzelnen angeführten Betrieben.

> *„Also der Schlachtbetrieb sagt einem wahrscheinlich nicht wirklich viel, wenn man das jetzt so liest. Vielleicht noch irgendwas, was den näher beschreibt (…) Weil es gibt ja solche und solche, denk ich.“ (I3, 23, Lehramt-Studentin)*

Ferner wurden von sechs Teilnehmer/-innen genauere Angaben auch zu den Haltungs- und Produktionsbedingungen gefordert. Im Zusammenhang hiermit wurde von den befragten Personen geäußert, dass sie gerne Informationen zum Einsatz von Antibiotikum oder Pestiziden hätten, und sie sich beispielsweise auch für das verwendete Futter in der Massentierhaltung interessieren würden.

Die Sicherung der Informationsqualität hat für den/die Verbraucher/-in einen hohen Stellenwert, zugleich wird die Qualitätskontrolle durch die Komplexität der Informationsketten erschwert. So wurde z. B. bei dem zufällig ausgewählten Testprodukt eines Rindfleisch Steaks bei den bereitgestellten Rezeptideen die mangelnde Informationsgüte angemerkt, da hier Rezepte zu Geflügel angezeigt wurden:

> *„Das ist aber doof, bei ‚weitere Rezepte‘ kommt dann Geflügel. Ja, das ist wahrscheinlich nicht ganz ausgereift.“ (I7, 59, Ingenieur)*

4.3 Nutzungskontexte und gesteigerte Attraktivität durch Mehrwertdienste

Für acht der befragten Personen spielten digitale Medien wie das Smartphone beim Einkauf von Lebensmitteln und auch beim anschließenden Konsum aktuell bereits eine Rolle. Diese gaben an, ihr Smartphone regelmäßig zum Notieren oder Abfotografieren von Einkaufslisten, zur Suche nach Rezepten oder neuen Produkten im Internet, oder während des Kochens als Timer zu nutzen und z. B. Kalorien und Nährstoffe mittels einer App zu tracken.

Da keiner der Befragten die *fTRACE* App vor dem Interview kannte, fokussierte unsere Befragung auf deren intendierter Nutzung. Hierbei gaben acht Teilnehmer/-innen an, die durch die App bereitgestellten Informationen hauptsächlich im Laden, vor dem Kauf nutzen zu wollen, um so eine bessere Kaufentscheidung treffen zu können:

> *„Direkt beim Kauf, also wenn ich in dem Laden bin und das Produkt in die Hand nehme, würde ich da schon gerne die Informationen sehn, weil ich dann entscheide ob ich's kaufe oder ob ich's nicht kaufe." (I8, 26, Einzelhandelskaufmann)*

Hierbei wurde von einem Befragten auch gewünscht, durch die App eine Kaufempfehlung zu erhalten. Diese sollte die persönlichen Präferenzen des/der Verbrauchers/-in berücksichtigen und könnte beispielsweise in Form einer Ampel angezeigt werden.

> *„Und man könnte dann so 'ne Art Farbcode entwickeln, (…) wo man dann selber auch einstellen kann, wie bei codecheck auch, mir sind diese Kriterien wichtig und das Produkt stimmt mit den voreingestellten Kriterien überein, dann Grün." (I7, 59, Ingenieur)*

Während sich fünf Teilnehmer/-innen eine regelmäßige Nutzung vorstellen konnten, äußerten acht andere, die App eher gelegentlich nutzen zu wollen, wenn sie gerade Zeit dafür fänden. Von drei Personen wurde auch angegeben, die App nur in speziellen Einzelfällen nutzen zu wollen, wenn man ein Produkt eines Herstellers das erste Mal kauft:

> *„Aber ich würde mir das halt wahrscheinlich dann wirklich nur einmal ankucken und nicht für jedes Fleisch, weil das ist wirklich nur für mich, für die Information, einmal zu wissen, wie's tatsächlich abläuft, interessant und danach immer das gleiche." (I5, 21, Maschinenbau-Studentin)*

Von drei der Befragten wurde angeregt, die Attraktivität und damit die Akzeptanz zu steigern, indem man die Nutzung mit weiteren Mehrwerten kombiniert. Es wurde z. B. die Möglichkeit angesprochen, mit Hilfe eines entsprechenden Systems bei Rückrufaktionen betroffene Produkte schnellstmöglich zu identifizieren und die Verbraucher/-innen besser und schneller, noch vor dem Konsum selbiger, warnen zu können. Auch wurde von sechs Befragten die Angabe eines Rezepts als Zubereitungsvorschlag für das gescannte Produkt zwar als grundsätzlich gute Idee bewertet, allerdings war ihnen die Auswahl dabei zu gering, da lediglich eine Rezeptidee angezeigt wurde:

> *„Hier ist die Kategorie Rezepte aufgelistet. Find' ich allgemein sehr gut, weil man dann direkt 'ne Idee hat, was man mit dem Produkt zubereiten kann. Hier ist aber leider nur ein Rezept dargestellt, vielleicht könnte man da noch ein, zwei vielleicht*

dazu tun, um nochmal bisschen mehr Auswahl zu haben." (I8, 26, Einzelhandelskaufmann)

Von drei Verbraucher/-innen wurde diesbezüglich erwähnt, dass sie eher auf eigene Rezepte oder andere Quellen, z. B. *chefkoch.de* oder *Google*, zurückgreifen würden. Für diese Teilnehmer/-innen war die Angabe einer Rezeptidee deshalb weniger interessant, und könnte stattdessen durch eine Verlinkung mit anderen Informationssystemen für Verbraucher/-innen ergänzt oder ersetzt werden.

4.4 Vertrauen

In Abschnitt 2.1 wurde dargelegt, dass es sich bei der Rückverfolgbarkeit in weiten Teilen um Vertrauensinformationen handelt. Für die Akzeptanz solcher Systeme wie *fTRACE* ist es deshalb zentral, dass Verbraucher/-innen der Quelle der Informationen vertrauen, glaubwürdige und zuverlässige Angaben zu machen. In den Interviews zeigte sich jedoch, dass dreizehn der Teilnehmer/-innen der Lebensmittelindustrie generell misstrauisch gegenüberstehen:

„Wie ja das Wort Industrie schon sagt, eher schlecht. Also Industrie ist immer auf Gewinn optimiert und wenig auf vordergründige Qualität, wenn der Preis zum Tragen kommt. Also mein Vertrauen ist nicht so groß." (I12, 34, Logistikplaner)

Es zeigte sich, dass dieses generelle Misstrauen sich in vielfacher Hinsicht auch auf eine kritische Bewertung bereitgestellter Informationen übertrug. Diese geraten in Verdacht von Werbung, bei der die Glaubwürdigkeit und die Objektivität zweifelhaft ist:

„Auch ist es interessant, dass hier bei der Herstellung dann ein Video angeboten wird, wo's per Hand geschnitten wird, was ich definitiv für nicht glaubwürdig halte, dass da wirklich einer mit dem Messer sitzt." (I11, 25, Philosophie-Student)

„Es gibt hier ja auch die Kategorie Qualität, da ist jetzt nicht ersichtlich, wo die Informationen herstammen." (I6, 22, Studentin Kommunikationswissenschaften)

„Allerdings muss man jetzt sagen, das ist 'ne eigene Seite, von denen erstellt. Das (…) ist deren Meinung. (…) Das ist halt Werbung auf 'nem höheren Level." (I4, 23, Maschinenbau-Studentin)

„(…) irgendwie find' ich das kurios, dass (…) es kein Drittkontrolleur ist, sondern man selber macht das dann. Wenn man grade bei Vertrauen ist, irgendwie find' ich's grade 'nen bisschen fragwürdig vielleicht." (I11, 25, Philosophie-Student)

Trotz dieses generellen Misstrauens sahen neun Teilnehmer/-innen in Ansätzen wie *fTRACE* eine Chance, das Vertrauen der Verbraucher/-innen in die Qualität von Lebensmitteln und in die Lebensmittelindustrie zu erhöhen. Der größte Vorteil wurde in der Schaffung von mehr Transparenz gesehen, die Verbraucher/-innen helfe, die Produktionsprozesse besser nachzuvollziehen. Zudem sahen fünf der Befragten darin einen positiven Effekt, dass durch die bessere Dokumentation der Prozesse der Betrug mit Lebensmitteln zwar nicht vollständig verhindert, aber zumindest reduziert und die Verantwortlichen besser zur Rechenschafft gezogen werden können.

„Ja, das würde das Vertrauen bestimmt erhöhen. Ich find' da kann's nicht genug Transparenz geben eigentlich." (I9, 25, VWL-Student)

„Es würde besser werden, weil ich ja dann einfach 'nen Einblick hätte über die gesamte Lieferkette und somit auch 'nen gesteigertes Vertrauen." (I12, 34, Logistikplaner)

„Da [bei der Rückverfolgbarkeit] gibt's dann viel weniger Möglichkeiten zu tricksen. Wobei […] kriminelle Energie kann natürlich auch solche Prozesse betreffen, ja." (I7, 59, Ingenieur)

Insgesamt zeigt sich, dass das Vertrauen in Informationsquelle und die von ihr bereitgestellten Informationen ein kritischer Faktor für die Nutzerakzeptanz sind.

5 Diskussion

Die Rückverfolgbarkeit stellt eine wichtige Maßnahme zur Qualitätssicherheit in der Lebensmittelindustrie dar. Obwohl Verbraucher/-innen eine wichtige Zielgruppe darstellen, wurden sie in der Forschung zu Rückverfolgbarkeitssystemen bisher kaum betrachtet. Wenngleich unser Sample von 16 Verbraucher/-innen sowie der Fokus der Studie auf eine bestimmte App und ein bestimmtes Produkt keine repräsentativen Schlüsse erlaubt, liefert sie dennoch wichtige Erkenntnisse zu Bedürfnissen von Verbraucher/-innen und möglichen Nutzungshürden, die bei der Gestaltung von Verbraucherinformations-Apps berücksichtigt werden sollten.

Eine wichtige Erkenntnis unserer Studie ist, dass die Verbraucher/-innen aus unserem Sample der digitalen Rückverfolgbarkeit gerade von frischen Lebensmitteln generell positiv gegenüberstehen und bestehende Lösungen wie *fTRACE* prinzipiell begrüßen, diese ihnen aber bisher weitestgehend unbekannt sind und deshalb im Alltag nicht genutzt werden. Eine erste Maßnahme wäre daher die bessere Kommunikation der vorhandenen Lösungen.

Unsere Studie zeigt ferner, dass eine ansprechende Informationsvisualisierung für die Nutzerakzeptanz wichtig sein kann. Dabei können visuelle Elemente wie Bilder, Karten und Ampelsysteme helfen, die Informationen nutzergerecht zu kommunizieren. Wie unsere Ergebnisse erkennen lassen, kann die Informationsaufbereitung jedoch eine Gratwanderung sein. Sie sollte nüchtern, aber ansprechend gestaltet sein, damit beim Verbraucher oder der Verbraucherin nicht der Eindruck entsteht, dass es sich um Werbung handelt. Dabei hat unsere Studie gezeigt, dass vor allem der Aspekt des Vertrauens in die dargestellten Informationen zentral ist. Unsere Teilnehmer/-innen bringen der Lebensmittelindustrie insgesamt ein geringes Vertrauen entgegen, was auch das Vertrauen in die Informationen aus Rückverfolgbarkeits-Apps schwächt. Eine Zertifizierung der Rückverfolgungsinformationen durch unabhängige Stellen sowie digitale Signaturen und neue technische Ansätze (wie z. B. Blockchain zum Nachweis von Kühlketten, vgl. [35–37]) könnten dazu beitragen, das Vertrauen in die Korrektheit der dargestellten Informationen zu steigern. Zudem könnte die Einbindung von Informationen von Verbraucherzentralen und NGOs sowie Benutzerbewertungen der Informationen und Produkte das Vertrauen stärken.

Ein genereller Nachteil von App-Ansätzen wie *fTRACE* ist, dass sie eine aktive Suche nach Informationen voraussetzen und von der Initiative der Verbraucher/-innen abhängig sind. Deshalb ist zu vermuten, dass solche Apps von vielen Verbraucher/-innen nur sporadisch und in bestimmten Situationen genutzt werden – z. B. wenn Verbraucher/-innen ein neues Produkt von einem bisher unbekannten Hersteller kaufen, sich mit Muße „einfach mal" informieren wollen, oder durch einen neuen Bericht zu einem Lebensmittelskandal verunsichert sind. Jedoch sollte daraus nicht der Schluss gezogen werden, dass die digitale Rückverfolgbarkeit für den/die Verbraucher/-in unwichtig ist. Für die Nutzer/-innen aus unserem Sample scheint es vielmehr wichtig zu sein, sich bei Bedarf gezielt über ein Lebensmittel, seine Herstellung und seine Verarbeitung informieren zu können, was sicherlich besonders bei frischen Lebensmitteln eine Rolle spielt. Technisch gesehen sollte jedoch mittelfristig das aktive Einscannen durch die Bereitstellung eines digitalen Kassenzettels ergänzt werden, bei dem Rückverfolgbarkeitsinformationen integriert sind und so eine proaktive Informationsversorgung ermöglicht [28, 29] wird. Darüber hinaus haben wir gezeigt, dass die Nutzerakzeptanz durch die Einbindung weiterer Informationsdienste gesteigert werden kann, wie z. B. Rezeptvorschläge, Vergleichsangebote für nachhaltige Produkte, sowie die Möglichkeit, sich bei Rückrufaktionen oder beim Ablaufen des Verzehrdatums (VD) des Produkts automatisch benachrichtigen zu lassen.

Unsere Ergebnisse machen nicht zuletzt auch deutlich, dass der Gestaltung des Point-of-Sale eine zentrale Rolle zukommt. Hierbei kann der Lebensmitteleinzelhandel eine wichtige Funktion übernehmen, indem er prominenter auf *fTRACE*-Produkte aufmerksam macht und z. B. durch das Aufstellen interaktiver Public-Displays Kundinnen und Kunden erlaubt, Produkte ohne Zuhilfenahme des eigenen Smartphones einzuscannen und Informationen abzufragen. Hierdurch ließe sich zum einen der Bekanntheitsgrad von Rückverfolgbarkeitssystemen stark verbessern, zum anderen unterstützte man dadurch die Verbraucher/-innen dabei, informierte Kaufentscheidungen zu treffen.

Für die weitere Arbeit an dem Thema wäre es interessant, unsere qualitativen Ergebnisse durch repräsentative Befragungen und Langzeitstudien zur Nutzung von Apps und Diensten wie *fTRACE* zu ergänzen. Unser Artikel hat dazu einen wichtigen, ersten Beitrag geleistet.

Referenzen

1. Opara, L.U.: Traceability in agriculture and food supply chain. Journal of Food Agriculture and Environment. 1, 101–106 (2003).
2. Ringsberg, H.: Perspectives on food traceability: a systematic literature review. Supply Chain Management: An International Journal. 19, 558–576 (2014).
3. Regattieri, A. et al.: Traceability of food products: General framework and experimental evidence. Journal of food engineering. 81, 347–356 (2007).
4. Skilton, P.F., Robinson, J.L.: Traceability and normal accident theory. Journal of Supply Chain Management. 45, 40–53 (2009).
5. Commission, E.: Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Communities. 31, 1–24 (2002).
6. Sarig, Y.: Traceability of food products. (2003).
7. Abbott, H.: Managing product recall: A comprehensive guide to establishing a product recall plan. Pitman Pub. (1991).
8. Dabbene, F., Gay, P., Tortia, C.: Traceability issues in food supply chain management: A review. Biosystems engineering. 120, 65–80 (2014).
9. Karlsen, K.M. et al.: Literature review: Does a common theoretical framework to implement food traceability exist? Food control. 32, 409–417 (2013).
10. Karlsen, K.M. et al.: Critical criteria when implementing electronic chain traceability in a fish supply chain. Food Control. 22, 1339–1347 (2011).
11. Morrissey, M.T., Almonacid, S.: Rethinking technology transfer. Journal of food engineering. 67, 135–145 (2005).
12. Abad, E. et al: RFID smart tag for traceability and cold chain monitoring of foods. Journal of food engineering. 93, 394–399 (2009).
13. GS1: Rückverfolgbarkeit in der Lieferkette: Grundlagen und Prozesse. GS1 (2018).
14. Thakur, M., Sørensen, C.-F., Bjørnson, F.O., Forås, E., Hurburgh, C.R.: Managing food traceability information using EPCIS framework. Journal of Food Engineering. 103, 417–433 (2011).
15. Thakur, M., Forås, E.: EPCIS based online temperature monitoring and traceability in a cold meat chain. Computers and Electronics in Agriculture. 117, 22–30 (2015).
16. Schoenheit, I.: Die volkswirtschaftliche Bedeutung der Verbraucherinformation. Politikfeld Verbraucherschutz. Beiträge einer Veranstaltungsreihe, Potsdam. 47–64 (2004).
17. Clement, R.: Internet-Ökonomie. Springer Berlin Heidelberg, Berlin, Heidelberg (2016).
18. Campos, S., Doxey, J., Hammond, D.: Nutrition labels on pre-packaged foods: a systematic review. Public health nutrition. 14, 1496–1506 (2011).
19. Szykman, L.R., Bloom, P.N., Levy, A.S.: A proposed model of the use of package claims and nutrition labels. Journal of Public Policy & Marketing. 228–241 (1997).
20. Hagen, K.: Nährwertkennzeichnung: Die Ampel erreicht die Verbraucher am besten. Wochenbericht des DIW Berlin. 22, 2–12 (2010).
21. Zühlsdorf, A., Spiller, A.: Trends in der Lebensmittelvermarktung. Göttingen: Agrifood Consulting GmbH. (2012).
22. Strünck, C.: Die Verbraucherpolitik braucht Pragmatismus statt wirklichkeitsferner Leitbilder. Wirtschaftsdienst. 91, 165–168 (2011).
23. Gregori, D. et al.: Evaluating food front-of-pack labelling. International journal of food sciences and nutrition. 65, 177–186 (2014).

24. Lammenett, E.: Preissuchmaschinen und Vergleichsportale. In: Praxiswissen Online-Marketing. pp. 263–267. Springer (2017).
25. Evizal, A.K., Rahim, S.K.A., Rahman, T.A., Rosa, S.L.: Traceability Software for the Food Industry. In: Advances in Food Traceability Techniques and Technologies. pp. 191–206. Elsevier (2016).
26. Gonzalez-Miranda, S. et al.: Future supermarket: overcoming food awareness challenges. In: Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2013 Seventh International Conference on. pp. 483–488. IEEE (2013).
27. Karpischek, S., Michahelles, F.: my2cents—Digitizing consumer opinions and comments about retail products. In: Internet of Things (IOT), 2010. pp. 1–7. IEEE (2010).
28. Stevens, G., Maslim, D., Placke, A., Blum, D.: eBon: Eine Designstudie zum Verbraucherschutz 2.0. Mensch & Computer 2014-Online Workshopband. (2014).
29. Bossauer, P., Hanschke, S., Stevens, G.: Mehrwerte auf Basis digitaler Kassenzettel: eine verbraucherinformatische Studie. In: Nachhaltiges Wirtschaften im digitalen Zeitalter. pp. 135–150. Springer (2018).
30. Ozarslan, S., Eren, P.E.: MobileCDP: A mobile framework for the consumer decision process. Information Systems Frontiers. 1–22 (2015).
31. Scholten, H., Verdouw, C.N., Beulens, A., van der Vorst, J.: Defining and analyzing traceability systems in food supply chains. In: Advances in Food Traceability Techniques and Technologies. pp. 9–33. Elsevier (2016).
32. Kurz, A., Stockhammer, C., Fuchs, S., Meinhard, D.: Das problemzentrierte Interview. In: Qualitative Marktforschung. pp. 463–475. Springer (2009).
33. Rohde, M. et al.: Towards a paradigmatic shift in IS: designing for social practice. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. p. 15. ACM (2009).
34. Charmaz, K., Belgrave, L.L.: Grounded theory. The Blackwell encyclopedia of sociology. (2007).
35. Zhou, Z., Zhou, Z.: Application of internet of things in agriculture products supply chain management. In: Control Engineering and Communication Technology (ICCECT), 2012 International Conference on. pp. 259–261. IEEE (2012).
36. Tian, F.: A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things. In: Service Systems and Service Management (ICSSSM), 2017 International Conference on. pp. 1–6. IEEE (2017).
37. Schütte, J., Fridgen, G. et al.: Blockchain und Smart Contracts: Technologien, Forschungsfragen und Anwendungen. (2017).

Umweltbewusstsein durch audiovisuelles Content Marketing? Eine experimentelle Untersuchung zur Konsumentenbewertung nachhaltiger Smartphones

Danielle Warnecke[1], Felix Redepenning[1], Frank Teuteberg[1]

[1] Universität Osnabrück, Unternehmensrechnung und Wirtschaftsinformatik, Osnabrück, Deutschland.
{danielle.warnecke,fredepenning,frank.teuteberg}@uni-osnabrueck.de

Abstract.
Informations- und Kommunikationstechnologie (IKT) hat heute einen hohen Stellenwert in unserer Gesellschaft, wobei insbesondere Smartphones sich mit immer kürzeren Produktlebenszyklen zur Lifestyle-Komponente entwickeln. Dabei werden sowohl im Herstellungsprozess als auch bei der Kaufentscheidung vorrangig ökonomische Entscheidungskriterien betrachtet, wodurch die sozio-ökologischen Folgen für Mensch und Umwelt an Produktions- und Entsorgungsorten vernachlässigt werden. Der Markt für nachhaltige Produkte leidet zudem unter erheblichen Informationsineffizienzen. Transparente Kommunikation zwischen nachhaltig agierenden Unternehmen und verantwortungsbewussten Konsumenten kann somit als essenziell für die Kaufentscheidung in diesem Bereich angesehen werden. Eine systematische Literatursuche und das gewählte Fallbeispiel des Fairphone2 bilden die Basis der experimentellen Untersuchung eines entwickelten Content Marketing Videos, dessen zugrundeliegendes Hypothesenmodell innerhalb einer Online-Befragung mit 284 Teilnehmenden evaluiert wird. Der vermutete Effekt auf die Konsumentensensibilität gegenüber Nachhaltigkeitsaspekten konnte in der Erhebung beobachtet werden und bietet somit eine Orientierung für weitere Forschungsansätze zur Werbewirkung im Bereich nachhaltiger IKT.

Keywords: Nachhaltigkeitsmarketing, Content Marketing, nachhaltige Smartphones, audiovisuelles Marketing.

1 Einleitung

Nachhaltigkeit rückt immer mehr in den Fokus politischen und gesellschaftlichen Handelns und sensibilisiert eine breite Öffentlichkeit für soziale, ethische und ökologische Aspekte [1]. Insbesondere im Bereich der Unterhaltungselektronik steht der Konsument als Nachfrager im Fokus nachhaltigen Entscheidens [2]. Dass Hersteller von Smartphones sich nur vereinzelt an Nachhaltigkeitsaspekten bei der Produktgestaltung orientieren, zeigen aktuelle Studien und Berichterstattung zu den Arbeits- und Produktionsbedingungen [3,4] sowie dem Ressourcenverbrauch bei der Herstel-

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

lung [5,6]. Trotzdem sind die Verkaufszahlen für Smartphones in den letzten Jahren stark angestiegen [7], was auch auf die immer kürzer werdenden Produktlebenszyklen zurückzuführen ist [8], sodass ein Smartphone im Schnitt bereits nach zwei Jahren gegen ein neueres Modell ausgetauscht wird [9].

Die stärksten Belastungen entstehen bei der Rohstoffgewinnung und Altgeräteentsorgung durch Kontakt mit toxischen Mineralien und Verbrennungsdämpfen [9,10]. Aufgrund dieser gefährlichen Arbeitsbedingungen und mangelhafter Ausrüstung kommt es hierbei zu erheblichen Schäden für Mensch und Umwelt [9,10].

Hersteller wie z. B. Fairphone B. V. oder Shift GmbH setzen mit ihrer Produkt- und Firmenphilosophie an dieser Stelle an und entwickeln Smartphones nach nachhaltigen Gesichtspunkten [11,12]. Allerdings orientierten sich bisherige Kampagnen des Nachhaltigkeitsmarketings häufig an Kampagnen für konventionelle Güter, obwohl sich die Wege zur Kaufentscheidung (auch Customer Journey) bei nachhaltig orientierten Konsumenten stark von weniger Umweltbewussten unterscheiden [13,14,15].

Der Nutzen von nachhaltigen Produkten für Konsumenten resultiert aus der Übereinstimmung persönlicher Werte mit den Markenwerten und einer Gewissensentlastung [16]. Es bestehen jedoch gravierende Informationsineffizienzen, sodass nachhaltige Produkte oftmals vom Verbraucher nicht als solche wahrgenommen werden [16]. Transparente Kommunikation zwischen Unternehmen und Konsumenten kann somit als entscheidend für den Erfolg nachhaltiger Produkte angesehen werden [18]. Im Rahmen des Content Marketings werden Tatsachen ansprechend aufbereitet und zu Werbezwecken genutzt [19]. Insbesondere audiovisuelle Kommunikation wirkt dabei sehr einprägsam [20].

Die diesem Beitrag zugrundeliegende Forschungsfrage lautet daher: Kann Konsumentenaufklärung in Form eines Content Marketing Videos für nachhaltige Smartphones zu einem verbesserten Umweltbewusstsein und somit zu einer erhöhten Kaufbereitschaft gegenüber nachhaltigen Smartphones beitragen? Zur Beantwortung dieser Frage wird zunächst anhand von Literatur und dem Fallbeispiel Fairphone 2 ein solches Video konzipiert und umgesetzt. Es wird als Stimulus in einem Online-Experiment (N=195) eingesetzt und entsprechend der entwickelten Hypothesen auf seine Wirksamkeit hin untersucht. Die Ergebnisse werden abschließend aus wissenschaftlicher und wirtschaftlicher Perspektive diskutiert.

2 Forschungsdesign

2.1 Methodisches Vorgehen

Es wird ein deduktiver Forschungsansatz in Anlehnung an [21] gewählt, bei dem Theoriebildung und Evaluation auf Basis wissenschaftlicher Literatur und Beobachtungswerten erfolgen. Die einzelnen Prozessschritte sind in Abb. 1 dargestellt.

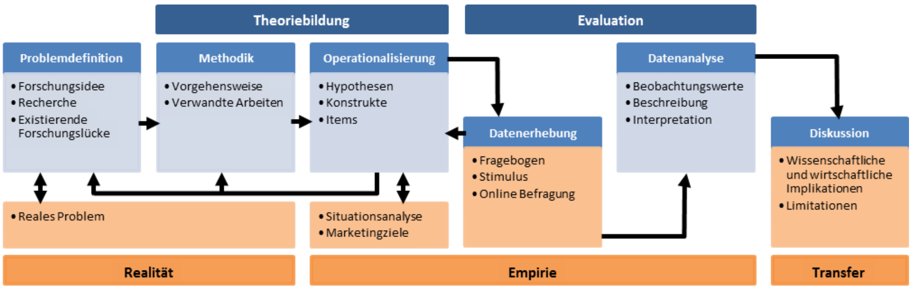


Abbildung 1. Deduktiver Forschungsprozesses modifiziert nach [21]

Nach der Problemdefinition und Skizzierung der Forschungsidee dient die Vorrecherche zur Überprüfung der Forschungslücke und zur Identifikation verwandter Arbeiten (siehe 2.2). Da durch diesen Teil der Untersuchung keine vollständige Beantwortung der Forschungsfrage ermöglicht wurde, wird im nächsten Schritt ein Fallbeispiel zur konkreten Betrachtung ausgewählt. Hierzu wird die Ausgangssituation des gewählten Unternehmens in Form einer Fallstudie systematisch analysiert, wobei kaufrelevante Faktoren, wie z. B. Unternehmensphilosophie, umweltschonende Ressourcen und Langlebigkeit des Smartphones berücksichtigt werden [22]. Die Ergebnisse der Situationsanalyse (siehe 3.1) dienen anschließend als Ausgangspunkt der formulierten Marketingziele (3.2) und der Entwicklung eines Content Marketing Videos (3.3).

Zur Evaluation dieses Lösungsansatzes werden die Marketingziele in ein Hypothesenmodell überführt (4.1) und anhand standardisierter Items operationalisiert (4.2). Neben den Items ist das Video als Stimulus zentraler Bestandteil des durchgeführten Online Experiments. Die Datenerhebung und -analyse (siehe Kap. 5) liefern die Basis zur anschließenden Diskussion (siehe Kap. 6) von Implikationen und Limitationen der Forschungsarbeit.

2.2 Verwandte Arbeiten

Es wurde eine systematische Literatursuche nach [23] in wissenschaftlichen Datenbanken[1] anhand der Suchbegriffe: *„Content Marketing, Nachhaltigkeitsmarketing, Marketing, Smartphone, Green IT, nachhaltig** und *IKT“* durchgeführt, wobei deutsche und englische Synonyme in verschiedenen Kombinationen verwendet wurden. Hierbei wurden 16 verwandte Arbeiten identifiziert, die die Bereiche IKT und Nachhaltigkeit oder Marketing und Nachhaltigkeit verbanden.

Die identifizierte Literatur des Nachhaltigkeitsmarketings lässt sich in drei Themengebiete einteilen, wobei keine dieser Arbeiten einen konkreten IKT-Bezug aufweist: Nachhaltigkeitsmarketing als Teildisziplin des Marketings, wobei vordergründig Eigenschaften und Entstehungsgeschichte behandelt werden [18,24], der Beitrag des Nachhaltigkeitsmarketings zur Lösung sozio-ökologischer Probleme [25] und

[1] EBSCO host, Science Direct, WISO-Net, AIS Electronic Library, Springer Link.

Reaktionen von Konsumenten auf Nachhaltigkeit sowie unterschiedliche Kategorien nachhaltig orientierter Konsumenten [26,27].

Mit Bezug zu nachhaltiger IKT haben [28], ein umfassendes Lehrbuch, das durch die Entwicklung nachhaltiger IKT gekennzeichnet ist und unterschiedliche Fallstudien zu nachhaltiger IKT zusammenfasst, veröffentlicht. [29] tragen die Vorteile nachhaltiger IKT aus Unternehmenssicht zusammen, während [9] und [30] explizit die Nachhaltigkeit von Mobiltelefonen untersuchen, wobei keine dieser Arbeiten Aspekte der Wissensvermittlung oder Kaufentscheidung durch Marketingmaßnahmen berücksichtigt.

Die größten Parallelen zur vorliegenden Untersuchung weist die Arbeit von [31] auf. Es wird untersucht, inwiefern nachhaltige Produkte von einer speziellen Marketingstrategie profitieren können. Jedoch wird hier der Einfluss klassischer Marketingmethoden auf die Konsumentenakzeptanz von Biokraftstoffen behandelt, die einer völlig anderen Produktkategorie als Smartphones angehören [32]. Demgegenüber wurde innerhalb des „B2B Technology Content Survey Report 2014" herausgestellt, dass Content Marketing (CM) die Kaufentscheidung auch für erklärungsbedürftige Produkte positiv beeinflussen kann [19,33]. Anhand der Arbeiten von [34], [35] und [36] wird insbesondere der Nutzen von CM bei der Informationsvermittlung hervorgehoben. Diese Eigenschaften werden im hier vorliegenden Beitrag genutzt, um den nötigen Informationstransfer zu unterstützen.

3 Audiovisuelles Content Marketing als experimenteller Stimulus

Das Ziel des zu entwickelnden Stimulus ist es, das Umweltbewusstsein und somit die Kaufbereitschaft für ein nachhaltiges Smartphone bei Konsumenten durch Wissenstransfer signifikant zu erhöhen. Als Fallbeispiel wird hierzu das niederländische Unternehmen Fairphone B. V. herangezogen, da dieses mit dem Fairphone 2 das aktuell bekannteste nachhaltige Smartphone vertreibt [37]. Im ersten Schritt wird hierzu die Situation des Unternehmens mit Hilfe der SWOT-Methode analysiert, woraus dann die Marketingziele abgeleitet werden, die als Rahmen zur Gestaltung des audiovisuellen Reizes in Form eines Videos dienen.

3.1 Situationsanalyse von Fairphone

Für die Situationsanalyse wird die SWOT-Methode genutzt, wobei Stärken, Schwächen, Chancen und Risiken des Unternehmens identifiziert werden. Eine externe Analyse betrachtet zunächst Chancen und Risiken, eine anschließende interne Analyse fokussiert Stärken und Schwächen des Unternehmens [22]. Die Ergebnisse der Analyse (siehe Tab.1) dienen als Ausgangspunkt für die Entwicklung von Marketingzielen [38].

Tabelle 1. Ergebnisse der SWOT-Analyse

Interne Analyse	
Stärken	Schwächen
• Fähigkeit zur Verbesserung von Funktionalität (Fairphone 1 zu Fairphone 2) • Mitarbeiterloyalität • Große Medienpräsenz • Guter Ruf	• Geringe produzierte Stückzahl • Geringer Umsatz • Geringes Unternehmensbudget • Wenige mögliche Zulieferer • Höhere Kosten (durch faire Rohstoffe)
Externe Analyse	
Chancen	Risiken
• Produktlanglebigkeit (mind. 5 Jahre) • Zusammenarbeit mit fairen Minen • Recyclingmaterialen verstärkt einsetzen • Gute Arbeitsbedingungen für Mitarbeiter an Produktionsorten • Recyclingprogramm für Altgeräte	• Geringe technische Performance • Preis ab 530 € • Tests und Rankings berücksichtigen selten sozio-ökologische Kriterien, sodass die Stärken des Fairphone 2 nicht honoriert werden.

Die identifizierten *Stärken* des Unternehmens liegen vor allem in der starken Unternehmensphilosophie [39], die auch durch Mitarbeiter und Medien wiedergegeben wird [37]. Die *Schwächen* hingegen liegen in geringem Produktionsvolumen, Umsatz und Kapitalausstattung im Branchenvergleich [40]. Zum einen limitiert die Unternehmenspolitik größtmöglicher Nachhaltigkeit und Transparenz die Auswahl potentieller Lieferanten [41,42], zum anderen besitzt Fairphone als Hersteller dieses Volumens geringe Verhandlungsmacht [43].

Die *Chancen* leiten sich aus den nachhaltigen Unternehmenszielen ab. Als primäres Unternehmensziel verfolgt Fairphone ein langlebiges Produktdesign, wobei die Lebensdauer eines Fairphone 2 auf fünf Jahre ausgerichtet ist [44]. Ein weiteres Ziel des Unternehmens ist die intensivere Zusammenarbeit mit fairen Minen und die vorrangige Verwendung recycelter oder erneuerbaren Ressourcen, außerdem sollen die Arbeitsbedingungen für Mitarbeiter an den Produktionsorten verbessert werden [44]. Schließlich fördert Fairphone durch ein eigenes Recyclingprogramm die Wiederverwendung alter Smartphones und die Weiterverwertung ihrer Komponenten [45].

Als *Risiken* zu betrachten sind der hohe Innovationsdruck der Smartphone-Branche [8], die vergleichsweise geringe Performance und der Preis (ab 530€) des Fairphones [46]. Fairphone sieht sich nicht als ein wettbewerbsorientiertes Unternehmen, sondern möchte stattdessen „als Inspiration für Kunden und andere Unternehmen der Industrie" dienen [47].

3.2 Marketingziele

Aus Mangel an Bewertungsnormen für nachhaltige IKT wurde häufig „Green Washing" durch falsche oder geschönte Nachhaltigkeitsberichte und Ökobilanzen praktiziert [15,48]. Aus diesem Grund sollten Werbebotschaften im Rahmen des Nachhaltigkeitsmarketings Konsumenten mit umfassenden Informationen versorgen und dabei vordergründig Tatsachen vermitteln [9,15,49]. Konsumenten könnten somit für die Notwendigkeit von nachhaltigen Smartphones sensibilisiert werden, sodass deren Wertschätzung (Zahlungsbereitschaft) für nachhaltige Rohstoffe, Arbeitsbedingungen und Kooperationen steigt [50,51]. Daraus lassen sich drei Hauptziele für die Unternehmenskommunikation ableiten:

- Wissen über sozio-ökologische Probleme in Minen und Produktionsstätten vermitteln,
- Wissen über nachhaltige Ansatzpunkte des Fairphone 2 vermitteln,
- Inhalte wahrheitsgemäß und transparent kommunizieren, um Vertrauen zu schaffen.

Besonders einprägsam wirken audiovisuelle Elemente der Unternehmenskommunikation [19]. Daher wird im Abschnitt 3.3 ein Video konzipiert, das über die Folgen herkömmlicher Smartphones und die Vorteile des Fairphone 2 anhand von Tatsachen im Sinne des Content Marketings aufklärt.

3.3 Audiovisuelles Konzept und Design

Das Video „Aus dem Leben eines Smartphones" wurde entsprechend der genannten Marketingziele anhand realer Produkt- und Unternehmenseigenschaften von Fairphone umgesetzt (abrufbar unter: http://bit.ly/audioCM (Anhang A)).

Nach einer kurzen thematischen Einführung werden die sozio- ökologischen Probleme genannt, die während der Rohstoffgewinnung und -herstellung, der Produktion, der Nutzung und dem Recycling von Smartphones entstehen. Im Anschluss werden Fairphones Lösungsansätze für jede der einzelnen Lebenszyklusphasen genannt und visualisiert. Die Animation mit gesprochenem Text und Hintergrundmusik hat eine Gesamtlänge von vier Minuten. Um ein Vertrautheitsgefühl zu erzeugen, werden die Zusehenden konsequent geduzt. Bei der Formulierung der Unternehmensziele wird weiterhin darauf geachtet „Wir-Botschaften" zu senden, um den Community-Gedanken von Fairphone auszudrücken. Drei Zitate der bekannten Film- und Serienfiguren Barney Stinson („Herausforderung angenommen!"), Meister Yoda („Tue es oder tue es nicht. Es gibt kein Versuchen.") und Remus Lupin („Es ist der Wert der Überzeugung, der den Erfolg ausmacht. Nicht die Anzahl der Anhänger.") motivieren zum vollständigen Anschauen des Videos.

4 Konzeption und Operationalisierung des Experiments

4.1 Konstrukte und Hypothesen gemäß Marketingzielen

Ausgangspunkt des Modellentwurfs sind die aus dem Fallbeispiel abgeleiteten Marketingziele [52]. Diese werden den Konstrukten „Wissen über Umweltprobleme", „Wissen über Nachhaltigkeitsvorteile des Produkts", „Vertrauen zum Unternehmen" und „Kaufbereitschaft für ein nachhaltiges Smartphone" wie in Tab. 2 zu sehen, zugeordnet und die vermuteten Zusammenhänge als überprüfbare Hypothesen umformuliert.
Die im Hypothesenmodell vermuteten Zusammenhänge werden in Abschnitt 4.2 anhand standardisierter Items in einem Fragebogen zusammengefasst.

Tabelle 2. Gegenüberstellung der Marketingziele und Konstrukte mit Hypothesen

Marketingziel	*Hypothesen je Konstrukt*	
Wissen über sozio-ökologische Probleme in Minen und Produktionsstätten vermitteln	*Wissen über Umweltprobleme (WüU)* H1: Informativ gestaltete Aufklärungsvideos beeinflussen das Wissen über Umweltprobleme beim Betrachter positiv. H2: Je größer das WüU ist, desto größer wird auch die Kaufbereitschaft für ein nachhaltiges Smartphone sein.	*Kaufbereitschaft für ein nachhaltiges Smartphone*
Wissen über nachhaltige Ansatzpunkte des Fairphone 2 vermitteln.	*Wissen über Nachhaltigkeitsvorteile des Produkts (WüN)* H3: Objektiv informativ gestaltete Aufklärungsvideos beeinflussen das Wissen über Nachhaltigkeitsvorteile des Produkts beim Betrachter positiv. H4: Je größer das WüN ist, desto größer wird auch die Kaufbereitschaft für dieses nachhaltiges Smartphone ausfallen.	
Vertrauen schaffen durch wahre und transparente Inhalte	*Vertrauen zum Unternehmen (V)* H5: Objektiv informativ gestaltete Aufklärungsvideos beeinflussen das Vertrauen zum Unternehmen positiv. H6: Je größer das V ist, desto größer wird auch die Kaufbereitschaft für ein nachhaltiges Smartphone ausfallen.	

4.2 Fragebogenentwurf

Da beim vorliegenden Hypothesenmodell ausschließlich latente Variablen mit festgelegter Beziehungsrichtung vorliegen, handelt es sich um ein reflektives Messmodell [53]. Die Verwendung multipler Items bei reflektiven Messmodellen gilt bei einer Anzahl von mind. 3 Items pro Konstrukt als wissenschaftlich anerkannt [54], da somit Verzerrungen sowie zufällige Messfehler einzelner Indikatoren ausgeglichen und die Reliabilität erhöht werden können [53,55]. Außerdem werden, wie in Tab. 3 beispielhaft für WüU dargestellt, mehrere Indikatoren als Aussagen in Fragebatterien zusammengefasst, um eine bestmögliche Validität zu erreichen [55]. Es wurden angepasste Formulierungen verwendet und ein Wechsel zwischen positiv und negativ formulierten Items umgesetzt (vollständige Liste der Items in Anhang B), um unaufmerksame Probanden und inkonsistentes Antwortverhalten aufzudecken [53]. Die Erhebung erfolgt mittels einer fünfstufigen Likert-Skala von „stimme überhaupt nicht zu“ (1Punkt) bis „stimme voll zu“ (5 Punkte). Die Antworten der Teilnehmenden lassen sich zu einem Testwert addieren, der die Einstellung der Person gegenüber dem Konstrukt widerspiegelt [56].

Tabelle 3. Items des Konstrukts WüU

Chiffre	*Item*	*Quelle*
WüU1	Ich würde mich selbst als jemanden beschreiben, der Ahnung von den sozio- ökologischen Problemen eines Smartphones hat.	[57]
WüU2	Ich kann jemand anderem die Vorteile eines nachhaltigen Smartphones erklären.	[57]
WüU3	Das ökologische Gleichgewicht ist sehr empfindlich und leicht durcheinander zu bringen.	[58]
WüU4	Natürliche Ressourcen sollten für zukünftige Generationen bewahrt werden.	[59]
WüU5	Der Zustand unserer Umwelt kann unsere Gesundheit beeinflussen.	[59]

Der Fragebogen umfasst 35 Fragen entlang der vier Item-Batterien zu den Konstrukten WüU, WüN, V und KB, sozio-demographischen Angaben sowie Einleitung und Schlusswort. Zu Beginn des Fragebogens erfolgt eine Randomisierung der Teilnehmenden, sodass diese zufällig der Kontrollgruppe (KG) oder Experimentiergruppe (EG) zugewiesen werden. Im Anschluss an die Einleitung erhält die EG den audiovisuellen Stimulus. Für die EG wurde eine durchschnittliche Bearbeitungsdauer von 18 Minuten im Pre-Test (14 Teilnehmende) beobachtet.

5 Datenerhebung und -analyse des Experiments

Die Datenerhebung erfolgt mithilfe einer Online-Befragung, eines Computer Self-Administered Questionnaire (CSAQ). Es wird ein Posttest-Only Control Group Design angewandt, um die Wirksamkeit des Videos (Stimulus) durch die Gegenüberstellung von Experimentier- und Kontrollgruppe zu überprüfen [55]. Die Befragung wird online im Zeitraum von April bis Juli 2018 per LimeSurvey durchgeführt, wobei die Probanden via Facebook, Whatsapp und E-Mail-Verteiler zur Teilnahme aufgefordert werden. Von 284 Teilnehmenden wurden unvollständig und inkonsistent beantwortete Fragebögen eliminiert, sodass der vollständig bereinigte Datensatz für die statistische Analyse 195 Fälle umfasst. Die statistische Auswertung der erhobenen Daten erfolgt mithilfe der Statistik- und Analyse-Software SPSS von IBM. Hypothesen werden bis zum Zehn-Prozent Signifikanzniveau angenommen.

5.1 Deskriptive Statistik

Das Convenience Sample setzt sich aus 58% männlichen und 41% weibliche Teilnehmenden zusammen. 89% besitzen als höchsten Schulabschluss das Fachabitur oder besser, wobei 50% der Teilnehmer zwischen 20 und 29 Jahren sind. 49% der Smartphones nutzen iOS, 48% Android und 3% sonstige Betriebssysteme. Laut der Probanden sind 56% dieser Smartphones neuer als 2 Jahre, wobei das älteste 10 Jahre alt ist. 10% der Geräte wurden im Dezember 2016 gekauft und 30% kosteten 499€ oder mehr. Hierin zeigt sich eine überraschend hohe Zahlungsbereitschaft, wobei nur

5% angaben, ihr Smartphone nach sozio-ökologischen Gesichtspunkten ausgewählt zu haben.

Die beiden Items „Natürliche Ressourcen sollten für zukünftige Generationen bewahrt werden." und „Der Zustand unserer Umwelt kann unsere Gesundheit beeinflussen." des Konstrukts WüU wurden mit durchschnittlichen Werten von 4,7 und 4,8 sehr hoch bewertet. Der Durchschnitt über alle fünf Items lag für WüU bei 3,9.

Das Konstrukt WüN wurde im Durchschnitt mit 3,9 bewertet, wobei die Items „Die Nutzung nachhaltiger Smartphones ist ein guter Ansatz, um den verschwenderischen Umgang mit natürlichen Ressourcen zu reduzieren." und „Die Nutzung eines nachhaltigen Smartphones ist ein guter Ansatz, um natürliche Ressourcen zu schützen." am höchsten bewertet wurden.

Den niedrigsten Durchschnittswert von 3,4 erreichte V, wobei die Aussage „Es besteht die Möglichkeit, dass die Nutzung eines nachhaltigen Smartphones der Umwelt schadet." mit 3,7 (nach Invertierung der Skala) am stärksten abgelehnt wurde.

5.2 Inferenzstatistik

Im Folgenden werden die erhobenen Testwerte der Konstrukte WüU, WüN, V und KB abhängig von verschiedenen Gruppierungsvariablen mithilfe des nichtparametrischen Mann-Whitney-U-Tests miteinander verglichen. Der MWU-Test prüft, ob eine signifikante Differenz zwischen den Mittelwerten zweier Verteilungen besteht. Dazu werden Rangreihenfolgen gebildet, indem die Testwerte aufsteigend angeordnet und die durchschnittlichen Ränge von Gruppen miteinander verglichen werden [60,61]. Die Nullhypothese des MWU-Tests besagt, dass es keinen Unterschied zwischen den abhängigen Variablen gibt [60]. Es wird somit überprüft, ob sich die Werte beider Gruppen signifikant voneinander unterscheiden. Dadurch kann eine Aussage über den Einfluss der jeweiligen Gruppierungsvariablen getroffen werden. Zuerst wird der Unterschied zwischen der KG und EG in der Konstruktbewertung gemessen, wobei die Werte zeigen, ob die Nullhypothese des Tests verworfen werden kann. Für das erste Konstrukt WüU kann die Nullhypothese verworfen werden, seit $p = 0{,}01 < \alpha = 0{,}1$. Für das zweite Konstrukt WüN kann die Nullhypothese ebenfalls verworfen werden, seit $p = 0{,}07 < \alpha = 0{,}1$. Für das dritte Konstrukt V kann die Nullhypothese nicht verworfen werden, da seit $p = 0{,}39 > \alpha = 0{,}1$ ist. Die Ränge der Experiment- und Kontrollgruppe unterscheiden sich für die Testwerte dieses Konstrukts folglich nicht.

Mithilfe des MWU-Tests werden außerdem die Durchschnittsränge der drei erstellten Testwerte für die Gruppe mit und ohne Kaufbereitschaft für ein nachhaltiges Smartphone miteinander verglichen. Dabei weist sich anhand der p-Werte ein signifikanter Zusammenhang zwischen WüU und KB ($p = .00$) aus. Es weist sich jedoch kein signifikanter Zusammenhang zwischen WüN und KB ($p = .20$) sowie zwischen V und KB ($p = .24$) aus. Die Nullhypothese wird hier demnach für zwei der drei Konstrukte beibehalten (siehe Abb.2).

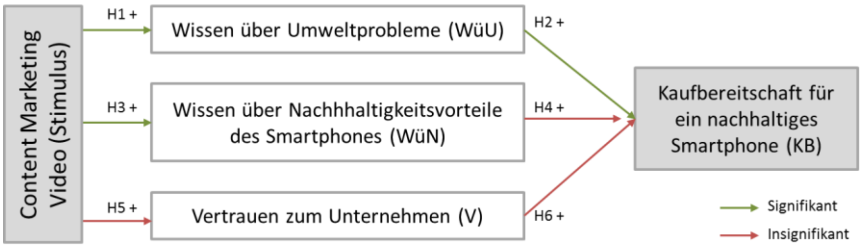


Abbildung 2. Hypothesenmodell nach Überprüfung

Weitere mögliche Zusammenhänge zwischen den sozio-demographischen Merkmalen und der KB werden mithilfe des Chi-Quadrat (χ^2) -Tests untersucht. Dieser ebenfalls nichtparametrische Test prüft, ob die beobachtete Verteilung der Variablen mit der erwarteten Verteilung übereinstimmt. Die Nullhypothese des χ^2-Tests besagt, dass die Variablen statistisch unabhängig sind bzw. kein Zusammenhang zwischen ihnen besteht [60]. Der Anpassungsgütetest beweist zunächst einen signifikanten Unterschied der Häufigkeiten zwischen erhöhter KB und niedriger KB (χ^2 = 74,06; df = 1; p = 0,00). Das zeigt, dass mehr Befragungspersonen eine niedrige KB aufweisen. Es zeigt sich ebenfalls eine signifikante Beziehung zwischen dem Geschlecht und der Kaufbereitschaft (χ^2 = 3.47; df = 1; p = 0,06). Frauen weisen demnach eine höhere KB als Männer auf. Eine signifikante Beziehung zwischen Alter und der KB konnte nicht identifiziert werden (χ^2 = 4,50; df = 5; p =0,48). Darüber hinaus ergibt sich eine signifikante Beziehung zwischen KB und Fachhochschulabschluss (χ^2 = 2.86; df = 1; p = 0,09) und zwischen KB und Universitätsabschluss (χ^2 = 3.26; df = 1; p =0,07).

6 Diskussion

Orientierungshilfen für den Konsumenten im Bereich nachhaltiger Produkte mit Aufklärungsfunktion wie das Biosiegel in der Lebensmittelindustrie [62] sowie das Fairtrade Siegel in der Lebensmittel- und Textilindustrie, das für sozial nachhaltig gehandelte Waren steht, sind bereits gesellschaftlich etabliert [63]. Siegel dieser Art kommen jedoch im Bereich IKT bislang nur vereinzelt zum Einsatz[32].

Da sich neben signalgebenden Labels besonders visuelle und audiovisuelle Medien zur produktbezogenen Wissensvermittlung eignen [34], geht der hier gewählte Experimentstimulus über reine Symbolwirkung hinaus und klärt den potenziellen Konsumenten über die Nachhaltigkeitsaspekte des beworbenen Smartphones (Fairphone2) in Form eines animierten Videos auf. Der Stil des Aufklärungsvideos folgt einem Mix aus emotionaler Ansprache (Aufmerksamkeit halten) und objektiver Informationsvermittlung (Mehrwert), wobei die persönliche Ansprache per Du und Wir-Formulierungen die Fairphone-Philosophie widerspiegelt. Das WüU und das WüN konnten laut Probandenbewertung dadurch positiv beeinflusst werden, wobei 16 von 18 freiwilligen Kommentaren der Befragungspersonen zu dem Video zustimmend ausfallen (siehe Anhang C). Der vermutete positive Effekt kann folglich in Form der

H1 (Stimulus beeinflusst das WüU positiv) und H3 (Stimulus beeinflusst das WüN positiv) durch die beobachteten Werte jeweils bestätigt werden. Hierbei werden insbesondere die Schäden an Menschen und Umwelt als Folge des Konsums konventioneller Geräte stärker (negativ) bewertet als die alleinige Hervorhebung sozio-ökologischer Vorteile nachhaltiger Smartphones (positiv). Außerdem werden Erklärungen anhand der Produktlebensphasen und der jeweiligen unternehmerischen Ansätze, wie z.B. Metalle aus fairen Minen als Rohstoff für die Herstellung, modulare Bauweise bei der Produktentwicklung, Reparaturanleitungen auf der Webseite, Rücknahme und Recyclingangebote für Altgeräte, etc. positiv bewertet. Jedoch konnte kein Effekt auf den Vertrauensaufbau beobachtet werden (Ablehnung H5). Die sachliche, weniger emotionale und einmalige Kontaktaufnahme ist für den Vertrauensaufbau demnach als ungeeignet einzustufen, denn sie verbessert das Vertrauen zum Unternehmen nicht.

Da in dieser Untersuchung das WüU, die ein Smartphone verursachen kann im Zusammenhang mit der gemessenen KB für dieses steht, kann außerdem ein positiver Effekt des Videos auf die Kaufbereitschaft beobachtet und somit die H2 angenommen werden. Ein Zusammenhang zwischen dem WüN eines Produktes und der KB sowie V und der KB konnte hingegen nicht nachgewiesen werden. H4 und H6 werden demzufolge abgelehnt. Dieses Ergebnis deutet daraufhin, dass die Kommunikation einer nachhaltigen Unternehmensphilosophie als noch wichtiger einzustufen ist als die transparente Kommunikation nachhaltiger Produkteigenschaften.

Die Auswertung der sozio-demographischen Merkmale ergab weiterhin, dass Universitätsabsolventen einen signifikant höheren WüU-Durchschnittsrang erreichen. Anhand der durchgeführten Chi-Quadrat Tests konnte außerdem nachgewiesen werden, dass Absolventen einer Universität oder einer Fachhochschule eine signifikant höhere KB für ein nachhaltiges Smartphone besitzen. Die Chi-Quadrat Tests zeigten weiterhin, dass Frauen eine höhere KB als Männer aufweisen. Ein statistisch zuverlässiger Zusammenhang zwischen Alter und der KB konnte hingegen in dieser Untersuchung nicht beobachtet werden. Daraus ergibt sich zunächst das Bild einer weiblichen Akademikerin, die Ziel von CM nachhaltiger Smartphones sein könnte.

Limitierend auf dieses Ergebnis wirkt sich aus, dass die beobachtete Kaufbereitschaft insbesondere bei sozial erwünschten Produkten nicht zwingend der realen Kaufentscheidung entspricht [64]. Jedoch lässt sich die Aussagekraft dieser Variable durch eine sehr konkrete Produktvorstellung (Eigenschaften des Fairphone2) erhöhen und sich somit mindestens als Näherungswert heranziehen [65]. Die erzielte Stichprobe ist ferner als nicht vollständig repräsentativ einzustufen, da 50% der Befragungspersonen zwischen 20 und 29 Jahren alt sind und 50% der Probanden einen Universitätsabschluss haben. Die Stichprobe stellt kein Abbild der Grundgesamtheit dar, jedoch bildet dieser Personenkreis einen wesentlichen Bestandteil des LOHAS[2] und somit der Zielgruppe für nachhaltige Smartphones [66].

Letztlich konnte durch das Experiment gezeigt werden, dass audiovisuelles Content Marketing zu einem besseren Umweltbewusstsein in Form von Wissen über

[2] LOHAS (Lifestyle of Health and Sustainability): Dieser Lifestyle verbindet Nachhaltigkeit, Verantwortung und Gemeinschaft mit Gesundheit, Wellness und Genuss [66].

Umweltprobleme und dem Wissen über die Nachhaltigkeitsvorteile von Smartphones beim Konsumenten signifikant beitragen kann, wohingegen ein Vertrauensaufbau durch einmalige Ansprache nicht zu beobachten war und die angegebene Kaufbereitschaft als Näherungswert, nicht als Faktum, betrachtet werden sollte.

7 Fazit

Aufbauend auf den Ergebnissen der Literaturanalyse werden anhand des Fallbeispiels Fairphone 2 Marketingziele hergeleitet, die für die theoretische Betrachtung in Hypothesen und Konstrukte überführt werden und anwendungsorientiert als Grundlage zur Entwicklung eines Content Marketing Videos dienen. Die Hypothesen werden im Rahmen eines Experiments per Online-Befragung mit 195 Teilnehmenden, bei der das entwickelte Aufklärungsvideo als Stimulus dient, überprüft. Der vermutete Effekt des Videos auf WüU und WüN und somit indirekt auf die Kaufbereitschaft nachhaltiger Smartphones konnte in der Untersuchung festgestellt werden, während ein Einfluss auf das Vertrauen zum Unternehmen durch das Video nicht beobachtet wurde. Content Marketing, insbesondere in Form eines Aufklärungsvideos ist folglich geeignet, um Wissen zu vermitteln. Ist dieses Wissen zielführend und zielgruppengerecht aufbereitet, kann es bei entsprechender Unternehmensphilosophie und Transparenz positiven Einfluss auf die Kaufbereitschaft des Konsumenten nehmen. Um das Zielgruppenprofil weiter zu spezifizieren kann zukünftige Forschung z. B. durch Forenanalyse der Fairphone-Community oder vertiefende Fallstudien in Unternehmenskooperation mit Fairphone, Shift GmbH oder anderen Anbietern zu weiterführenden Erkenntnissen beitragen.

Anhang

Der Anhang ist abrufbar unter folgender URL: http://bit.ly/audioCM.

Danksagung

Diese Arbeit ist Teil des Projekts "Nachhaltiger Konsum von Informations- und Kommunikationstechnologie in der digitalen Gesellschaft − Dialog und Transformation durch offene Innovation". Das Projekt wird vom Ministerium für Wissenschaft und Kultur des Landes Niedersachsen und der VolkswagenStiftung aus Landesmitteln des Niedersächsischen Vorab gefördert (Projektnummer VWZN3037).

Literaturverzeichnis

1. Statistisches Bundesamt: Nachhaltige Entwicklung in Deutschland: Indikatorenbericht 2016. Statistisches Bundesamt (2017)

2. Stanszus, M.: Mit der grünen Suchmaschine WeGreen bewusst finden und nachhaltig konsumieren: Die Macht der Konsumenten und mit welchen Werkzeugen der Attitude-Behaviour-Gap geschlossen werden kann. Zeitschrift für Wirtschafts- und Unternehmensethik. Bd. 13, Ausg. 3, 345-349 (2012)
3. Statista: Absatz von Apple iPhones weltweit vom 3. Geschäftsquartal 2007 bis zum 2. Geschäftsquartal 2016. http://de.statista.com (Zugegriffen: 15.09.2017).
4. Apple Inc.: Supplier Responsibility 2015 Progress Report. http://www.apple.com (Zugegriffen: 15.09.2017).
5. Bafilemba, F., Mueller, T., Lezhev, S.: The Impact of Dodd-Frank and Conflict Minerals Reforms on Eastern Congo's Conflict, https://enoughproject.org (Zugegriffen: 15.09.2017).
6. Amnesty International: "This is what we die for" Human rights abuses in the Democratic Rebublic of Congo power the global trade in cobalt. https://www.amnesty.org (Zugegriffen: 15.09.2017).
7. Business Wire: Absatz von Smartphones weltweit in den Jahren 2009 bis 2017 (in Millionen Stück). https://de.statista.com (Zugegriffen: 08.03.2018).
8. Kumar, A.P., Kannegala, S.S.: Green Devices and Hardware. In: Harnessing Green IT, S.23–38. Chichester (2012)
9. Cramer, B.W.: Man's need or man's greed: The human rights ramifications of green ICTs. Telematics and Informatics, vol. 29, no. 4, S. 337–347 (2012)
10. Murugesan, S., Gangadharan, G.R.: Green IT: An Overview. In: Harnessing Green IT, S.1–22. John Wiley & Sons, Ltd, Chichester (2012)
11. FairPhone B.V.: About Us. https://www.fairphone.com (Zugegriffen: 05.03.2018).
12. Shift GmbH: Shift Report 2016. https://www.shiftphones.com (Zugegriffen: 18.08.2018)
13. McDonald, S., Oates, C.J., Alevizou, P.J., Young, C.W., Hwang, K.: Individual strategies for sustainable consumption. Journal of Marketing Management, vol. 28, no. 3–4, 445–468 (2012)
14. Oates, C.J., McDonald, S., Alevizou, P.J., Young, C.W., Hwang, K., McMorland, L-A: Marketing Sustainability: Use of information sources and degrees of voluntary simplicity. Journal of Marketing Communications, vol. 14, no. 5, 351–365 (2008)
15. Peattie, K.: Golden goose or wild goose? The hunt for the green consumer. Business Strategy and the Environment, vol. 10, no. 4, 187–199 (2001).
16. Meffert, H., Rauch, C., Lepp, H.L.: Sustainable Branding - mehr als ein neues Schlagwort?!. Marketing Review St. Gallen, vol. 27, 28–35 (2010)
17. Galarraga Gallastegui, I.: The use of eco-labels: A review of the literature. European Environment, vol. 12, 316–331 (2002)
18. Belz, F.-M., Peattie, K.: Sustainability Marketing. John Wiley & Sons, Ltd, Chichester (2009)
19. Hilker, C.: Content Marketing in der Praxis. Ein Leitfaden - Strategie, Konzepte und Praxisbeispiele für B2B- und B2C-Unternehmen. Springer Fachmedien, Wiesbaden (2017)
20. Kreilaus C.: Web-Videos — Social Branding und Performance-Optimierung. In: Schulten M., Mertens A., Horx A. (eds) Social Branding. Gabler Verlag, Wiesbaden (2012)
21. Rogge, K.-E.: Methodenatlas. Springer-Verlag, (2013)
22. Hill, C.W.L., Jones, G.R.: The Strategic Management Process. Strategic Management- An Integrated Approach, 1–34. Houghton Mifflin Company, Boston (2004)
23. Vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: ECIS 2009 Proceedings, vol. 161, 2206–2217 (2009)

24. Polonsky, M.J.: An Introduction to Green Marketing. Electronic Green Journal, vol. 1, no. 2 (1994)
25. Kinoti, M.W.: Green marketing Intervention Strategies and Sustainable Development : A Conceptual Paper. International Journal of Business and Social Science, vol. 2, no. 23, 263–273 (2011)
26. Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E.A., Nelson, L.D.: Commitment and Behavior Change: Evidence from the Field. Journal of Consumer Research, vol. 39, 1070–1084 (2013)
27. McDonald, S., Oates, C.J., Thyne, M., Alevizou, P.J., McMorland, L.-A.: Comparing Sustainable Consumption Patterns Across Product Sectors. International Journal of Consumer Studies, vol. 33, no. 2, 137–145 (2009)
28. Dahlem, M., Erek, D.K., Glau, T., Henseler, D.R., Kusiak, K., Löser, F., Opitz, N., Rekers, J., Samulat, P., Schmidt, D.N.-H., Schilling, D.T., Stelzer, T.: Green IT - Erkenntnisse und Best Practices aus Fallstudien. In: Zarnekow, R., Kolbe, L. (Hrsg). Springer-Verlag, Berlin Heidelberg (2013)
29. Banerjee, S., Sing, T.Y., Chowdhury, A.R., Anwar, H.: Motivations to adopt Green ICT: A tale of two organizations. International Journal of Green Computing (IJGC), vol. 4, no. 2, 1–11 (2013)
30. Walser, A.: Mobiltelefone im Spannungsfeld von sozial-ökologischen Problemen und Kundenbedürfnissen. In: Nachhaltigkeits-Marketing in Theorie und Praxis, 211–226. Deutscher Universitäts-Verlag, Wiesbaden (2005)
31. Bernigau, S.: Eine Marketing-Strategie für nachhaltigere Biokraftstoffe in Deutschland: Ein Ansatz zur Verbesserung der Konsumentenakzeptanz?. Springer Gabler, Wiesbaden (2017)
32. Atkinson, L., Rosenthal, S.: Signaling the Green Sell: The Influence of Eco-Label Source, Argument Specificity and Product Involvement on Consumer Trust. Journal of Advertising, 43:1, 33-45 (2014)
33. Schach, A.: Advertorial, Blogbeitrag, Content-Strategie & Co. Springer Fachmedien, Wiesbaden (2015)
34. Harad, K.C.: Content Marketing Strategies to Educate and Entertain. Journal of Financial Planning, vol. 26, no. March 2013, 18–20 (2013)
35. Pophal, L.: How to create a content strategy. E Content, vol. 39, no. 2 (2016)
36. Steinbach, J., Krisch, M., Harguth, H.: Helpvertising. Springer Fachmedien, Wiesbaden (2015)
37. Cook, G., Elizabeth, J.: Guide to Greener Electronics. Washington, D.C. (2017)
38. Meffert, H., Christoph, B., Manfred, K.: Marketing - Grundlagen marktorientierter Unternehmensführung. 12. Auflage. Springer Fachmedien, Wiesbaden (2015)
39. Pakalski, I.: Fairphone 2 kann vorbestellt werden. http://www.sueddeutsche.de (Zugegriffen: 21.06.2018)
40. FairPhone B.V.: Bedrijsprofiel. https://www.fairphone.com (Zugegriffen: 05.03.2018)
41. FairPhone B.V.: List of Suppliers for the Fairphone 2. https://www.fairphone.com (Zugegriffen: 13.03.2018)
42. FairPhone B.V.: How We Work. https://www.fairphone.com (Zugegriffen: 06.03. 2018)
43. FairPhone B.V.: Kostenaufschlüsselung für das Fairphone 2. https://www.fairphone.com (Zugegriffen: 13.02. 2018)
44. FairPhone B.V.: Bei uns geht die Lebensdauer von Produkten in die Verlängerung. https://www.fairphone.com (Zugegriffen: 24.03. 2018)
45. FairPhone B.V.: Goals. https://www.fairphone.com (Zugegriffen: 06.03. 2018)

46. CHIP Digital GmbH: Smartphone Test: Handy Bestenliste. http://www.chip.de (Zugegriffen: 11.03.2018)
47. Schwarz, A.: Wir wollen die Industrie inspirieren. http://www.taz.de (Zugegriffen: 11.05.2018)
48. Padel, S., Foster, C.: Exploring the gap between attitudes and behaviour: Understanding why consumers buy or do not buy organic food. British Food Journal, vol. 107, no. 8, 606–625 (2012)
49. Carrigan, M., Attalla, A.: The myth of the ethical consumer – do ethics matter in purchase behaviour?. Journal of Consumer Marketing, vol. 18, no. 1, 560–578 (2001)
50. Hüser, A.: Institutionelle Regelungen und Marketinginstrumente zur Überwindung von Kaufbarrieren auf ökologischen Märkten. Zeitschrift für Betriebswirtschaftslehre, vol. 63, no. 3, 267–287 (1993)
51. Löffler, M.: Think Content!. 1. Auflage. Rheinwerk Verlag GmbH, Bonn (2014)
52. Herrmann, A., Homburg, C. (Hrsg.): Marktforschung. 2.Auflage. Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, Wiesbaden (2000)
53. Weiber, R., Mühlhaus, D.: Strukturgleichungsmodellierung. 2. Auflage. Springer Gabler, Berlin Heidelberg (2014)
54. Fuchs, C., Diamantopoulos, A.: Using single-item measures for construct measurement in management research: conceptual issues and application guidelines. Die Betriebswirtschaft, vol. 69, no. 2, 195–210 (2009)
55. Altobelli, C.F.: Marktforschung. UVK Verlagsgesellschaft mbH, Konstanz München (2011)
56. Greving, B.: Messen und Skalieren von Sachverhalten. In: Methodik der empirischen Forschung, 65–78. Springer Fachmedien, Wiesbaden (2009)
57. Crosno, J.L., Cui, A.P.: A Multilevel Analysis of the Adoption of Sustainable Technology. Journal of Marketing Theory and Practice, vol. 22, no. 2, 209–224 (2014)
58. Clark, C.F., Kotchen, M.J., Moore, M.R.: Internal and external influences on pro-environmental behavior: Participation in a green electricity program. Journal of Environmental Psychology, vol. 23, 237–246 (2003)
59. Haron, S.A., Paim, L., Yahaya, N.: Towards sustainable consumption: An examination of environmental knowledge among Malaysians. International Journal of Consumer Studies, vol. 29, no. 5, 426–436 (2005)
60. Burns, R.B., Burns, R.A.: Business Research Methods and Statistics Using SPSS. SAGE Publications Ltd, London (2013)
61. Janssen, J., Laatz, W.: Nicht parametrische Tests. In: Statistische Datenanalyse mit SPSS, 631–692. Springer Gabler, Berlin Heidelberg (2017)
62. Janssen, M., Hamm, U.: Product labelling in the market for organic food: Consumer preferences and willingness-to-pay for different organic certification logos. Food quality and preference 25.1, 9-22 (2012).
63. Loos, J., Bertels, V., Müller, S.: Die Wirkung des vertrauensstiftenden Fair Trade-Siegels auf die Kaufentscheidung von Verbrauchern. Macht des Vertrauens, Wiesbaden, 149-183 (2013)
64. Schnell, H., Hill, P.B.: Esser: Methoden der empirischen Sozialforschung. München 201.1 (2013).
65. Morwitz, V.G., Steckel, J.H., Gupta, A.: When do purchase intentions predict sales? International Journal of Forecasting 23.3, 347-364 (2007)
66. Glöckner, A., Balderjahn, I., Peyer, M.: Die LOHAS im Kontext der Sinus-Milieus; Marketing-Review St. Gallen, 27, 5, 36–41 (2010)

Towards Predictive Energy Management in Information Systems: A Research Proposal

David Heim[1], Gregor Friedrich-Baasner[1], Anna Fuchs[1], and Axel Winkelmann[1]

[1] University of Wuerzburg, Chair of Business Information Systems, Wuerzburg, Germany
{david.heim,gregor.friedrich-baasner,
a.fuchs,axel.winkelmann}@uni-wuerzburg.de

Abstract. The progressive energy transition, driven by the growing number of renewable energies, the increasing social importance of sustainable actions, as well as new technologies, causes major challenges for enterprises and power supply companies (PSCs). While the electricity price fluctuations will continue to increase in the future, the installation of smart meters and smart meter gateways is aimed to ensure grid stability. They provide the basis for communication between companies and PSCs. In order to make companies energy consumption predictable even before the energy is needed, an automated data exchange between an energy management system (EnMS) and enterprise resource planning (ERP) system is essential. Therefore, we address this problem by following five research steps to develop a prototype for predictive energy management in information systems.

Keywords: Energy transition, predictive energy management, information systems, demand-side management

1 Motivation and Current Developments

A predictive energy management system provides the foundation for a secure, eco-friendly and economical operation of the energy system. In the future, the produced amount of electricity should be consumed in order to exploit renewable energies and to avoid the current problem of buffer storage. The German Federal Government's goal is to cover 35 % of its electricity requirements by 2020 and 80 % by 2050 from renewable energies [1-3]. This leads to increased volatility on the electricity market [4], [5]. The far-reaching transition, driven by the liberalization of the energy market, the growing number of renewable supply, as well as increasing socially importance of sustainable actions and the change of paradigm (consumption must adapt to production) in the energy industry to decentralized feed-in causes major challenges in terms of information processing and business models [3], [6].

The resulting growing complexity of internal processes, increased customer requirements and rapidly changing economic conditions can be counteracted by the use of adequate information and communication systems [4]. Nowadays, ERP systems are an important component of companies' IT infrastructure. When implemented successfully, they support business processes by linking all areas of a

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

company and enable the integrated management of all enterprise resources [7]. This includes also production-relevant data such as orders, production planning and machine occupancy.

On the other hand, companies are increasingly using EnMS, which are tools used to control and monitor organizations energy flows. The ISO 50001 standard, published in June 2011, is the first international standard for EnMS and supports organizations in all industries in implementing such systems [8]. The main objective of such systems is to increase energy efficiency in order to achieve energy savings, thus optimizing energy consumption and saving costs [9]. With regard to production-relevant data, an EnMS provides information about the energy consumption of machines and production lines as well as their load profiles [10].

By combining the data from both systems, it is possible to calculate a predicted progression over time [4]. Based on this information, PSCs can use them to predict energy demand more precisely and, if necessary, to take appropriate countermeasures. Furthermore, companies can benefit from this development and optimize their production costs. At the same time, the enormous increase in electricity costs means that variable electricity rates can gain even more importance for companies [11], [12]. Significant cost savings potential can be generated even when production processes are postponed within the day. Due to Singhal and Swarup [13] energy prices can be highly volatile to systems conditions. While the magnitude of electricity price fluctuations will continue to increase in the future [14], companies can generate sustainable competitive advantages through intelligent production planning. Price fluctuations result in particular from the availability of renewable energies that cannot be planned precisely and are therefore dependent on external factors. Changes with regard to price fluctuations can be observed in particular in countries that are already increasingly relying on renewable energies [15-17]. An oversupply of electricity on the short-term energy market can result in negative prices. The number of hours during which negative electricity prices occur have increased significantly since 2010 [18]. Current research projects in the field of energy management, such as TechPlan[1], SynEnergie[2], FOREnergy[3], and PHI-Factory[4] underline the relevance of our research topic, but do not consider our approach in this way. An initial overview of the literature shows, that there are several publications on ERP-Systems and EnMS [4], [19-21]. However, the combination of both systems is currently not considered [22], [23].

2 Research Approach

The goal of our approach is to develop a prototype that allows an automated data exchange between ERP systems and EnMS. This should enable companies, to link their energy consumption with the current amount of available energy and the prices on the energy market. By aggregating this information across many companies, PSCs can more accurately predict demand and react if the amount of energy is not available at the moment. The aim of a bidirectional data exchange between companies and PSCs is to ensure grid stability and create incentive models for energy consumption.

This can be achieved, for example, by creating price incentives to shift the production of the companies to periods when demand of energy is lower.

In order to adequately address this problem, we divided our research approach into five steps, as shown in figure 1. The first step, we are working on, is to create an initial overview of current research projects, as mentioned in chapter 1, as well as several publications on ERP systems and EnMS. An exact breakdown of the scientific literature with the help of a structured literature review according to Webster & Watson [24] will be our second step. The focus of the review is on the differentiation between ERP systems and EnMS, as well as their functions. The literature and technology overview will represent our knowledge base as a basis for our iterative adjustments during the development of the prototype.

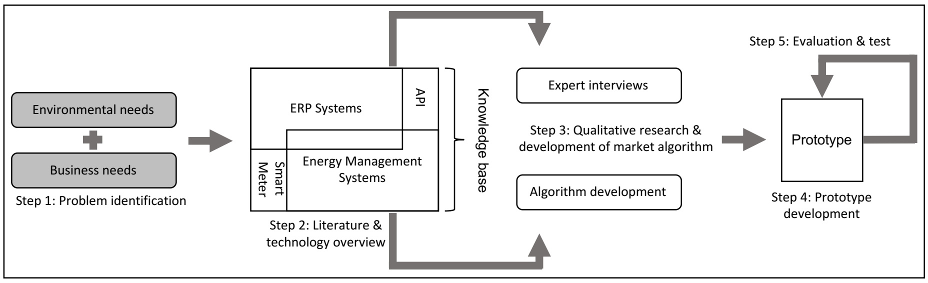


Figure 1. Research Approach

In order to adequately consider current issues and practical experiences, the requirements of companies for such a prototype have to be analyzed in the third step. To determine these, qualitative interviews with experts from different industries and company sizes are suitable [16]. Additionally, a matching algorithm should be developed, which can connect the data from the ERP system with the data from EnMS.

In the fourth step, this algorithm will be implemented into the existing infrastructures of the companies. To ensure a bidirectional data exchange between ERP systems and EnMS, the prototype contains standardized API's. This is required to enable the prototype to interact with both systems. The evaluation and test of the prototype will be an iterative process. In this case the evaluation is followed by further surveys through experts. The method of delphi studies is selected to gain these insights [25]. Ultimately, this project is intended to lead to a constantly improving tool, with a deep learning algorithm, that supports companies in their digitization.

3 Conclusion and Expected Contribution

Our research will provide the PSCs with more reliable data for their energy planning. However, the main condition, and therefore also a limitation, is that an ERP must be available at the company. At the same time, it aims to support companies in optimizing costs and enables new business models to be developed in the long term.

These new business areas are, for example, in the previously unknown topic of the energy broker. Furthermore, new services can also be created in the areas of billing, procurement and consulting [6]. Precisely, the supply fluctuations will arise from the increase in renewable energies, we assume that electricity price fluctuations will increase in line with market equilibria [15-17]. This offers both, companies and PSCs the opportunity to achieve competitive advantages through elaborated strategic action. However, these advantages can only be achieved on a reliable data basis. In our point of view, this is only possible on the basis of existing systems (ERP systems and EnMS) and the combination as well as the expansion of those. The ERP system already provides all necessary production data and the EnMS provides the corresponding energy data. By linking them, it is possible to adequately predict the required energy consumption. The linking of these systems leads to a better basis for decision-making and thus to greater insights. On the other hand, PSCs can use historical and weather data to calculate the amount of energy available and react to bottlenecks at an early stage. In this context, new technologies, like the smart meter can also be used in a meaningful way and contribute to the general objective of the energy turnaround. Here, smart meters and smart meter gateways provide the platform for the transmission of electricity prices as well as the respective demand levels. In addition, the solution to be developed will be established in familiar ecosystems, so that no transaction costs will arise for companies. One thing remains certain despite the change: Energy is the backbone of new innovations and will become an immanently important asset over the coming decades. In order to handle our planet carefully and to support the change in the energy industry, we saw ourselves induced to this research approach and have the noble goals to contribute to it. In our opinion, no major new system solution is required. Information systems have the necessary functions and the corresponding data foundation. The main objective here is to exploit the advantages of these mature systems by means of an adequate connection. This allows the potential of the systems to be expanded and enhanced. This research project is intended to test and evaluate this thought-provoking impulse.

References

1. Goebel, C., Jacobsen, H.-A., Razo, V., Doblander, C., Rivera, J., Ilg, J., Flath, C., Schmeck, H., Weinhardt, C., Pathmaperuma, D., Appelrath, H.-J., Sonnenschein, M., Lehnhoff, S., Kramer, O., Staake, T., Fleisch, E., Neumann, D., Strüker, J., Erek, K., Zarnekow, R., Ziekow, H., Lässig, J.: Energy Informatics. Current and Future Research Directions. Bus. Inf. Syst. Eng. 6, 25–31 (2013)
2. Gerpott, T.J., Paukert, M.: Determinants of Willingness to Pay for Smart Meters: An Empirical Analysis of Household Customers in Germany. Energy Policy. 61, 483–495 (2013)
3. Jagstaidt, U.C.C., Kossahl, J., Kolbe, L.M.: Smart Metering Information Management. Wirtschaftsinformatik. 53, 313–317 (2011)
4. Feuerriegel, S., Bodenbenner, P., Neumann, D.: Is More Information Better Than Less? Understanding the Impact of Demand Response Mechanisms in Energy Markets. In: Proc. 21st European Conference on Information Systems, pp. 1–12. (2013)

5. Strbac, G.: Demand Side Management: Benefits and Challenges. Energy Policy. 36, 4419–4426 (2008)
6. Doleski, O.D.: Geschäftsmodell – Methode zur Realisierung von Geschäftsideen für Utility 4.0. Springer-Verlag, Heidelberg (2016)
7. Becker, J., Kugeler, M., Rosemann, M.: Process Management A Guide for the Design of Business Processes. Springer-Verlag, Berlin (2014)
8. International Organization for Standardization: ISO 50001 Energy Management, https://www.iso.org/iso-50001-energy-management.html (Accessed: 31.10.2018)
9. Sequeira, H., Carreira, P., Goldschmidt, T., Vorst, P.: Energy Cloud: Real-Time Cloud-Native Energy Management System to Monitor and Analyze Energy Consumption in Multiple Industrial Sites. In: 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, pp. 529–534. (2014)
10. Schellong, W.: Analyse und Optimierung von Energieverbundsystemen. Springer-Verlag, Heidelberg (2016)
11. Grandel, M.: Das „Smart Metering Dilemma" – Strategische Überlegungen zum flächendeckenden Einsatz von Smart Metering. In: Servatius H.-G., Schneidewind U., Rohlfing D. (eds.) Smart Energy. pp. 221–231. Springer-Verlag, Heidelberg (2012)
12. Kunz, C., Müller, A., Saßning, D.: "Smart Grids" für die Stromversorgung der Zukunft. Optimale Verknüpfung von Stromerzeugern, -speichern und -verbrauchern. Renews Spez. 58 (2012)
13. Singhal, D., Swarup, K.S.: Electricity Price Forecasting Using Artificial Neural Networks. Int. J. Electr. Power Energy Syst. 33, 550–555 (2011)
14. Benini, M., Marracci, M., Pelacchi, P., Venturini, A.: Day-ahead Market Price Volatility Analysis in Deregulated Electricity Markets. In: IEEE Power Engineering Society Summer Meeting, pp. 1354–1359. IEEE Press, New York (2002)
15. Romanchenko, D., Odenberger, M., Göransson, L., Johnsson, F.: Impact of Electricity Price Fluctuations on the Operation of District Heating Systems: A Case Study of District Heating in Göteborg, Sweden. Appl. Energy. 204, 16–30 (2017)
16. Zareipour, H., Bhattacharya, K., Cañizares, C.A.: Electricity Market Price Volatility: The Case of Ontario. Energy Policy. 35, 4739–4748 (2007)
17. Lund, H., Münster, E.: Management of Surplus Electricity-production from a Fluctuating Renewable-energy Source. Appl. Energy. 76, 65–74 (2003)
18. Statista: Anzahl der Stunden mit negativen Strompreisen in Deutschland in den Jahren 2008 bis 2017, https://de.statista.com/statistik/daten/studie/618751/umfrage/anzahl-der-stunden-mit-negativen-strompreisen-in-deutschland/ (Accessed: 31.10.2018)
19. Curry, E., Hasan, S., O'Riain, S.: Enterprise Energy Management Using a Linked Dataspace for Energy Intelligence. In: Sustainable Internet and ICT for Sustainability (SustainIT), pp. 1–6. (2012)
20. Bunse, K., Vodicka, M., Schönsleben, P., Brülhart, M., Ernst, F.O.: Integrating Energy Efficiency Performance in Production Management – Gap Analysis Between Industrial Needs and Scientific Literature. J. Clean. Prod. 19, 667–679 (2011)
21. Vikhorev, K., Greenough, R., Brown, N.: An Advanced Energy Management Framework to Promote Energy Awareness. J. Clean. Prod. 43, 103–112 (2013)
22. Goldhahn, L., Bock, D., Eckardt, R., Weber, H.: Energetische Planungsalgorithmen für ERP-Systeme (EnPlan): Entwicklung Energetischer Planungsalgorithmen und deren Nutzbarmachung für den Einsatz in ERP-Systemen Metallverarbeitender KMU. ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetr. 107, 603–607 (2012)

23. Karnouskos, S., Colombo, A.W., Martinez Lastra, J.L., Popescu, C.: Towards the Energy Efficient Future Factory. In: 2009 7th IEEE International Conference on Industrial Informatics, pp. 367–371. IEEE Press, New York (2009)
24. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Q. 26, xiii–xxiii (2002)
25. Skinner, R., Nelson, R.R., Chin, W.W., Land, L.: The Delphi Method Research Strategy in studies of information systems. Commun. Assoc. Inf. Syst. 37, 31–63 (2015)

[1] TechPlan (http://www.techplan-erp.de/): TechPlan is a research project to improve the energy efficiency of SMEs in the metal industry. This should be achieved through energetic and technological planning of production processes.

[2] SynEnergie (https://www.fim-rc.de/forschung/forschungsprojekte/laufend/synergie/): The reserach project "SynEnergie" deals with synchronized and energy-adaptive production technologies for a flexible alignment of industrial processes and the fluctuating energy consumption.

[3] FOREnergy (http://forenergy.de/): The aim of the FOREnergy research project is the evaluation of concepts and solutions for an energy-flexible factory.

[4] PHI-Factory (http://phi-factory.de/): The goal of this research project is the development of technical and organisational solutions that enable industrial companies to save energy costs and support power grid stability at the same time.

A Web Browser-Based Application for Processing and Analyzing Material Flow Models using the MFCA Methodology

Christian Kunisch, Volker Wohlgemuth

HTW Berlin - University of Applied Sciences, Faculty 2: School of Engineering, Technology and Life, Berlin, Germany
Christian.Kunisch@htw-berlin.de, Volker.Wohlgemuth@htw-berlin.de

Abstract. In the context of IT for Green, the potential of the Material Flow Cost Accounting (MFCA) method has not yet been fully unleashed. Using the advantages of mobile data acquisition for material and energy flows, for instance right at a production plant, makes it easier to identify optimization potentials in sustainability management. Tapping this potential, for example through continuous data acquisition of sensor data from sensors in a production plant, can also be made more convenient. Studies conducted that the introduction and application of MFCA has brought significant ecological and economic benefits. They also revealed that there is very little MFCA specialized software on the market. This paper describes the objective and concept of a web browser-based application for the processing and analysis of material flow models and their evaluation with the MFCA methodology.

Keywords: MFCA, IT for Green, Web browser-based application, Angular, Material flow model.

1 Introduction

Reducing energy consumption, waste emissions and the use of primary resources helps to achieve sustainable development [1], which is why resource efficiency is a recent issue on the EU's policy agenda [1]. In the political orientation of the Federal Republic of Germany (German Resource Efficiency Programme II), too, the goal of resource-saving production is anchored as a field of action in the coming years [2]. Particularly the industrial sector should focus its current path towards more sustainable development, such as resource efficiency and cleaner production.

One way to increase material efficiency, i.e. to reduce the sum of input compared to output, is, among others, to uncover the true monetary value of waste generated in production processes through the methodology of material flow cost accounting (MFCA) and thus encourage entrepreneurs to recycle it or sell it to other entrepreneurs. This paper first briefly describes the current scientific status of the MFCA concept and then gives a detailed insight into the structure of the web browser-based application.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

2 Material Flow Cost Accounting

In the late 1990s MFCA has been developed in Germany [3-5], starting with first approaches of flow cost accounting [6]. In the following decade the Japanese Ministry of Trade and Industry (METI) strongly popularized MFCA in Japan [4]. To date, more than 300 Japanese companies have integrated material flow cost accounting [4].
Meanwhile MFCA is ISO-standardized to 14051:2011 (General framework) and 14052:2017 (Guidance for practical implementation in a supply chain). Now ISO 14053 is under development addressing MFCA in small and medium sized enterprises.

MFCA quantifies "the flows and stocks of materials within an organization in physical units (e.g. mass, volume) and the costs associated with those material flows are also evaluated. The resulting information can act as a motivator for organizations and managers to seek opportunities to simultaneously generate financial benefits and reduce adverse environmental impacts" [7].

The application of MFCA and the subsequent implementation of appropriate measures have shown that remarkable ecological and economic benefits can be achieved [4, 8]. A case study of an Argentinean sugar cane company proves that the use of MFCA and the resulting decision to acquire new facilities worth 350,000 US dollars saves 72,000 US dollars annually [9]. There is only a very limited number of software solutions specialized on MFCA, inter alia a desktop application „Umberto NXT MFCA" and "bw!MFCA" from the ifu Institut für Umweltinformatik Hamburg GmbH [10, 11] and the web application "Materialflusskostenrechner" of the VDI Zentrum Ressourceneffizienz GmbH [12].

Material flow cost accounting is to be promoted as a method because it optimally supports the combination of three perspectives. Firstly, the consideration of physical energy and material flows, secondly the monetary evaluation of savings potentials and thirdly the reduction potential of the climate impact. The material flow cost accounting method provides companies with information on waste and scrap losses. In particular, the true value of these losses is reported. This topic is therefore relevant for manufacturing SMEs, not least for reasons of increasing yields through increased resource efficiency. Based on achieved MFCA results, SMEs are enabled to optimize products and/or processes, to improve their business balance sheets and ultimately to contribute to securing Germany as a business location.

3 Web Browser-Based Application

Due to a lack of supporting software solutions for mobile data recording where the material and energy flows occur and for displaying the volume flows during tours of the production areas to support the development of process optimizations on site [13] the research project "MFCA mobile" was started in November 2017. The "Institut für Umweltinformatik Hamburg" (ifu) already developed a software for CO_2 balancing, material flow costing and visualization of material flows.

It also includes a material flow cost accounting component. But the software exists exclusively as a so-called desktop application for Microsoft Windows and can therefore only be used locally. A decentralized (mobile) data acquisition or an efficient on-site presentation of the collected data and models in the production areas is therefore not supported [13].

The central concept of the project is the combination of flexible data acquisition, validation and visualization on site with the opportunity of subsequent processing the data on the workstation computer using advanced tools that demonstrate optimization potential. One of the main objectives of the project is to implement a mobile variant to exchange data with the desktop version via a cloud service, creating new usage scenarios in the mobile context for material flow management and MFCA.

Through the mobile data acquisition and visualization and through the combination with an existing desktop application for MFCA, the application of the method for resource efficiency questions is significantly simplified. A further developed concept for data acquisition and representation of material flows will be developed and implemented. This should enable SMEs in the future to record their production processes in a structured way and thus implement operational material flow management. As an alternative to the autonomous approach, the efficient cooperation with experts/consultants will be optimally supported. Sankey diagrams are ideal for visualizing material and energy flows in production processes. They therefore also form a good basis for collecting material and energy flows as well as other data in production processes.

The interaction with a powerful desktop software should be used to intensify the cooperation between consultants and specialists in the company and to link the method of material flow cost accounting with an elegant data collection. The data collector in the company benefits from being able to check and modify the process modeling directly in the production environment on the tablet and to complete the model with data. The smooth and fast data exchange between consultant (desktop PC) and data supplier (tablet) in connection with the graphical representation of the production model leads to faster and better results. A special focus in the expansion of the concept will therefore be on the secure transmission of data by a cloud server including rights management. A commentary function for the work processes allows important work to be documented during the data acquisition tour in order to increase efficiency and to be noted for later implementation. The data to be entered i.e. can be:

- material and energy flows, the quantities of which are,
- Process specifications, their recipe factors, new inputs and outputs,
- Process parameters whose values are,
- Material properties, such as specific weight, price and more,
- Substance concentrations, substance properties and stocks.

The web browser-based application which is under development offers the following features:

- written in TypeScript with the frontend web application framework Angular,
- supports at least Chrome, Firefox and Edge,
- operable by mouse as well as by multitouch and
- is usable on smartphones, tablets and desktop.

The following Figure 1 shows an overview of the structure of the Web Browser-based application.

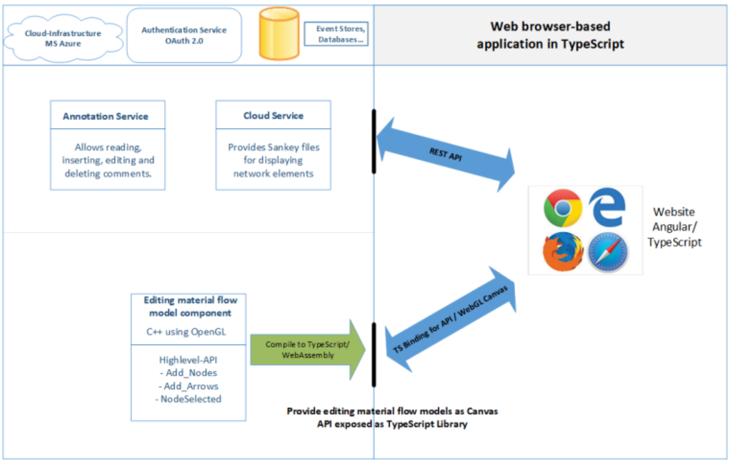


Figure 1. Overview of the structure of the web browser-based application

The Annotation-Service specifies metadata that is associated with web service implementations. It allows reading, inserting, editing and deleting of comments to the processes. The RESTful API is used for the communication between client and server [14]. Authentication and Authorization is implemented by the open protocol OAuth 2.0. It is the industry-standard protocol for authorization [15].

It allows secure authorization in a simple and standard method from web, mobile and desktop applications [16]. Other tools used in the infrastructure are GitLab Repositories [17], automated builds, deployment and test execution, a Build server at GitLab which consists of docker containers for building, deploying and testing and Microsoft Azure [18] for hosting the website whereat deployment can be started manually or automatically in GitLab.

4 Discussion and Outlook

The fact to gather data from mobile devices and push them into an existing material flow model offers the chance for an almost live calculation of the material flows and therefore the MFCA results too. This provides the opportunity to create a continuous and automated MFCA analysis. It is planned to have a mature prototype with these properties available in the first quarter of 2019. Completion is scheduled for January 2020. Currently we are implementing the authorization and authentication service for the web application. Further we are working on the routing and the router guards. With guards it is possible to add checks to restrict access to a user to certain pages [19].

This research project is funded by the "Deutsche Bundesstiftung Umwelt" (DBU). We would like to thank the DBU for their support.

References

1. EEA, European Environment Agency (2016). More from less – material resource efficiency in Europe. EEA report No 10/2016. Technical report, European Environment Agency.
2. Bundesregierung: Programm zur nachhaltigen Nutzung und zum Schutz der natürlichen Ressourcen, http://dip21.bundestag.de/dip21/btd/18/077/1807777.pdf (Accessed: 22.08.18)
3. Christ, K.L. & Burritt, R.L. (2014). Material flow cost accounting: a review and agenda for future research. Journal of Cleaner Production. http://dx.doi.org/10.1016/j.jclepro.2014.09.005 (Accessed: 22.08.18)
4. Schmidt, M. & Nakajima, M. (2013). Material Flow Cost Accounting as an Approach to Improve Resource Efficiency in Manufacturing Companies. Resources, 2, pp. 358-369. Doi:10.3390/resources2030358. ISSN 2079-9276.
5. Viere, T., Prox, M., Möller, A., Schmidt, M. (2011). Implications of Material Flow Cost Accounting for Life Cycle Engineering. Conference Paper. 18th CIRP International Conference on Life Cycle Engineering, Braunschweig.
6. Wagner, B. & Strobel, M. (1999). Kostenmanagement mit der Flusskostenrechnung. In Werkzeuge erfolgreichen Umweltmanagements; Freimann, J., Ed.; Gabler. Springer Fachmedien Wiesbaden, pp. 49–70.
7. ISO, International Organisation for Standardisation (2011). ISO 14051:2011. Environmental Management – Material Flow Cost Accounting – General Framework.
8. Schaltegger, S. & Zvezdov, D. (2015). Expanding material flow cost accounting. Framework, review and potentials. Journal of Cleaner Production, Vol. 108, pp. 1333-1341, http://dx.doi.org/10.1016/j.jclepro.2014.08.040.
9. Scavone, G.M. (2006). Challenges in internal environmental management reporting in Argentina. Journal of Cleaner Production, Vol 14, pp. 1276-1285.
10. ifu, Institut für Umweltinformatik Hamburg GmbH. Umberto NXT MFCA: https://www.ifu.com/umberto/mfca-materialflusskostenrechnung/, (Accessed: 23.08.18)
11. ifu, Institut für Umweltinformatik Hamburg GmbH. Umberto NXT MFCA: https://www.ifu.com/umberto/bwmfca/, (Accessed: 23.08.18)
12. VDI Zentrum Ressourceneffizienz GmbH. Materialflusskostenrechner: https://kostenrechner.ressource-deutschland.de/#/materialverluste (Accessed: 23.08.18)
13. Kunisch, C. & Wohlgemuth, V. (2018). Mobile Support of Material Flow Cost Accounting (MFCA) in Small and Medium Sized Enterprises. Conference Paper. 32nd EnviroInfo Conference. (Accepted for publication in Advances and New Trends in Environmental Informatics, Springer Nature Switzerland AG)
14. Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. Dissertation. University of California, Irvine. (Accessed: 21.11.18)
15. D. Hardt, Ed., Microsoft (2012). The OAuth 2.0 Authorization Framework. https://tools.ietf.org/html/rfc6749 (Accessed 29.08.18)
16. Parecki, A. (2018). OAuth 2.0: https://oauth.net/ (Accessed 29.08.18)
17. GitLab. https://about.gitlab.com/ (Accessed 29.08.18)
18. Microsoft Azure, Microsoft. https://azure.microsoft.com/de-de/ (Accessed 29.08.18)
19. Hussain, A., Daolrevo Ltd., Codecraft.tv, Router Guards: https://codecraft.tv/courses/angular/routing/router-guards/ (Accessed: 29.08.18)

Track 13:

Digital Work – Social, mobile, smart

Track Chairs:

Prof. Dr. Sven Laumer
Friedrich Alexander Universität Erlangen-Nürnberg

Dr. Christian Maier
Otto-Friedrich Universität Bamberg

Prof. Dr. Alexander Richter
IT University of Copenhagen

Prof. Dr. Matthias Trier
Universität Paderborn & Copenhagen Business School

On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work

Stephan Diederich[1], Alfred Benedikt Brendel[1], Lutz M. Kolbe[1]

[1] University of Goettingen, Chair of Information Management, Goettingen, Germany
stephan.diederich@stud.uni-goettingen.de,
{abrendel, lkolbe}@uni-goettingen.de

Abstract. Conversational agents (CA), i.e. software that interacts with its users through natural language, are becoming increasingly prevalent in everyday life as technological advances continue to significantly drive their capabilities. CA exhibit the potential to support and collaborate with humans in a multitude of tasks and can be used for innovation and automation across a variety of business functions, such as customer service or marketing and sales. Parallel to the increasing popularity in practice, IS researchers have engaged in studying a variety of aspects related to CA in the last few years, applying different research methods and producing different types of theories. In this paper, we review 36 studies to assess the status quo of CA research in IS, identify gaps regarding both the studied aspects as well as applied methods and theoretical approaches, and propose directions for future work in this research area.

Keywords: Conversational agent, virtual assistant, machine collaboration, literature review

1 Introduction

Conversational agents (CA), i.e. software that interacts with users via written or spoken natural language, increasingly permeate our lives. Nowadays, mobile devices are equipped with powerful agents by default, such as Siri or Google Assistant, offering support for a variety of tasks such as researching information, scheduling meetings, or sending messages. At its 2018 developer conference, Google demonstrated the potential of CA in the (near) future by showing their assistant autonomously making an appointment with a hairdresser in a live phone conversation [1]. For organizations, capable CA offer a variety of applications ranging from team collaboration, such as in the form of cognitive assistants in workshops [2], to service provision at the customer interface [3–5], and have attracted increasing interest in practical implementation in recent years [6]. CA have been around for several decades, starting with the agent ELIZA developed by Joseph Weizenbaum in 1966 which simulated a psychotherapist [7], yet several agents did not fulfill expectations in the past [4]. As most of these agents were primarily rule-based, their potential and capabilities were rather limited. However,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

driven by advances in natural language processing and machine learning, modern CA emerge into seemingly intelligent software that can be used to support various tasks and enhance human cognitive capabilities [8], hereby changing the allocation of task between humans and machines [9]. As CA become an integral part of our lives and application grows increasingly versatile, we believe there are multiple aspects to study ranging from the design of such artifacts, understanding emerging, collaborative work practices to the introduction of CA in organizations for automation and innovation.

While CA have been researched extensively in human-computer interaction (HCI) and computer science (CS) with the different foci prevalent in these disciplines, such as user trust in CA in HCI or optimizing natural language understanding in CS [4], they attracted the interest of the IS community recently [10] because of their profound impact on both organizational as well as individual level. As of now, a variety of IS studies address different aspects, such as design principles [4], information disclosure by the user [11, 12] or the impact of the CA on user's cognitive abilities [13]. Against this background, we aim to structure and analyze what is already known in IS research regarding CA and derive directions for future studies with the following research question:

RQ: What is the status quo and what are future directions for IS research on CA?

In order to address this question, we conduct a systematic review of IS literature with a focus on highly-regarded journals and conferences. We analyze 36 publications in terms of studied CA characteristics as well as the produced IS theory types and applied research methods. This allows us to better understand and structure existing work as well as to derive future research directions. We continue by providing the theoretical background of CA and outlining our research approach, a systematic, concept-centric review of IS literature [14]. We then present the insights from our analysis, discuss the results and close this paper by formulating research directions for future studies on CA.

2 Research Background

CA are systems that interact with people using natural language, thus simulating behavior of a human being [4, 15, 16]. Terms, such as virtual assistants, chatbots or dialogue systems are often used synonymously [4, 10]. As natural language can be written or spoken, CA communicate with their users via written language, speech, or both [17]. CA that communicate with users via written language, often referred to as chatbots, include agents such as Rose (a bot with the personality of a security analyst and hacker from San Francisco) or Mitsuku who can play games or reason with specific objects while simulating a 18-year old woman from Leeds. Examples in a company context are the CA of KLM, called BlueBot [18], that allows to find and book flights, the chatbot by H&M that provides personal shopping recommendations [10] or Amtrak's customer service bot that handles more than one million customer requests per year [19]. The most popular CA with speech-based input can be found on mobile devices that are used in our daily lives, such as Alexa, Siri, Microsoft's Cortana, or Google's Assistant.

Nowadays, more than 100.000 bots engage with users on a global scale on Facebook, and companies are showing great interest as more than 80% have already implemented bots for customer service or plan to do so by 2020 [6]. With several enterprise platforms in the market, such as IBM's Watson Conversation Service, the Microsoft Bot Platform or DialogFlow by Google, companies can now easily procure and customize off-the-shelf solutions to introduce CA in their organizations.

Due to the significant improvements of CA, the widespread diffusion of powerful mobile devices, and continuing trends toward digitization, a variety of CA has emerged in the past few years both in private as well as professional life. While general-purpose CA, such as Siri or Google Assistant cover a wide range of functions, domain-specific agents focus on specific tasks or domains. For example, CA can be used for automation in customer service [3, 4, 20], as digital sales assistants [21, 22] or in human resources to provide support for new employees [23]. Thus, in addition to the communication mode, CA can also be differentiated by the context in which they are used [4].

Furthermore, CA can be represented in different forms, such as with static virtual avatars [8, 24], interactive virtual avatars [25, 26], physical embodiment [27, 28] or be disembodied, i.e. lack any form of representation and only provide a natural language interface [29]. These different forms offer a variety of design options and provide non-verbal cues that influence how a CA is perceived by a user [30, 31]. For example, the experiments of Qiu and Benbasat [32] on the virtual static representation of a CA indicate that matching the ethnicity leads users to perceiving the CA as more enjoyable, sociable and useful. Additionally, Al-Natour et al. [33] investigated online shopping assistants with a focus on personality and behavioral similarity and found that customers evaluate CA that are similar to themselves more positively and show higher intentions to reuse them.

3 Research Approach

To review existing work on CA in IS research, we follow a literature review process based on the approaches of Webster and Watson [14], Brocke et al. [34], and Bandara et al. [35]. Our review consists of three phases. First, we gather studies regarding CA from established and relevant IS journals and conferences. Second, we code the identified literature along five dimensions: CA mode of communication, CA context, CA embodiment, IS theory type, and IS research method. Finally, we analyze the literature by using a concept matrix to assess the status quo of CA research and derive directions for future studies.

Table 1. Research approach phases

	Phase 1: Gather literature	**Phase 2: Code literature**	**Phase 3: Analyze literature**
Inputs	Publication databases and outlets	Literature database	Coded literature database
Methods	Literature search	Coding	Literature analysis
Steps	Conduct search and filter literature	Define coding dimensions and code literature	Create and interpret concept matrix
Results	Literature database	Coded literature database	Overview of CA research and implications

For the literature search process, the relevant literature outlets had to be identified. We decided to focus on highly-ranked publication outlets [36] and extend them by current studies from IS conferences. Hence, we focused on articles published in the basket of eight [37] and complemented these with selected conferences (ICIS, ECIS, WI, MKWI, HICSS, AMCIS, PACIS) to incorporate recent research because CA have just recently started to attract the attention of the IS community [10]. To gather publications, we used the Web of Science and AIS Electronic Library databases. Furthermore, we also included the website of the respective journal or conference if the outlet was not already included in the databases. The following search query was used:

Conversational Agent OR Virtual Assistant OR Dialogue System OR Chatbot

The full-text search was conducted in November 2018 by two of the authors. We omitted identical results and briefly scanned titles as well as abstracts to remove irrelevant articles. For example, studies with a focus on agents could research autonomous software agents that are used to fulfill specific tasks, e.g. in e-commerce, yet lack the conversational nature of the agents analyzed in this study. Similarly, the term agent included studies on human service agents, such as in customer service, which are not the focus of our review and thus were excluded. After this search and filtering process, 36 articles remained (see Table 2, outlets without relevant results are omitted).

Table 2. Literature search results

Journals	Total Found	Filtered
Journal of Management Information Systems	90	3
Journal of the Association for Information Systems	75	2
International Conference on Information Systems	383	13
European Conference on Information Systems	287	2
Pacific Asia Conference on Information Systems	198	4
Hawaii Int. Conference on System Sciences	100	5
Americas Conference on Information Systems	518	7
	Total	36

In phase two of our work we coded the literature along different dimensions. To assess the status quo, we chose a combination of dimensions related to the content of the studies (communication mode, context, embodiment) as well as regarding the methodological approach (research method) and type of produced knowledge (theory type). Thus, we were able to identify research gaps both related to the content as well as from a methodological and theoretical perspective. Difficult decisions during the coding process, such as assigning a research method in a study that uses a combination of methods, were discussed by both authors. For example, the study by Seeger et al. [38] develops a design framework and evaluates it based on an online experiment. In this case, we assigned the method "framework/conceptual model" as this can be considered the main contribution of the study. Furthermore, we differentiated the theory types "explanation" and "design and action" by the explicit formulation of design principles and patterns [39] whereas studies that explain user interaction with a CA often also provide valuable implications for the design (e.g. [40]).

Regarding the content-related dimensions, we determined the communication mode of the CA (text-based, speech-based, or both) to account for the fact that natural language can be written or spoken [17]. Furthermore, we assessed whether the study deals with a general-purpose CA or an agent that is domain-specific, i.e. used for a specific task or function [4]. Finally, we considered the embodiment of the CA [41], i.e. whether it has a static virtual representation, an interactive virtual representation, a physical representation or no form of representation at all. We extended these dimensions by the research methods and produced IS theory types. Palvia et al. [42, 43] assessed the research methods specifically in IS, thus we selected their methods to code our studies. With regard to the IS theory types, Gregor [44] distinguishes five types, which we used in our coding: Analysis (theory that describes and analyzes reality without the identification and structuring of cause-effect relations), explanation (theory that provides explanations for cause-effect relations but does not formulate propositions), prediction (theory that contains testable propositions but does not provide justification or explanation), explanation and prediction (theory that provides both causal explanation and testable propositions), as well as design and action (theory that informs the development of artifacts). In total, we coded the studies along five dimensions (Table 3) and all studies were assigned one characteristic per dimension.

In the third and final phase, we analyzed the coded literature by means of a concept matrix [14]. A concept matrix helps to view literature from a concept-centric [45] position and thus fosters an understanding of research beyond descriptive content summarization [14]. Furthermore, it helps to study the distribution of characteristics within the dimensions and paves the way for answering our research questions.

Table 3. Coding dimensions

<table>
<tr><th>Dimension</th><th colspan="10">Characteristics</th></tr>
<tr><td>CA communication mode</td><td colspan="3">Text-based</td><td colspan="4">Speech-based</td><td colspan="3">Both</td></tr>
<tr><td>CA context</td><td colspan="5">General-purpose</td><td colspan="5">Domain-specific</td></tr>
<tr><td>CA embodiment</td><td colspan="2">None</td><td colspan="3">Virtual static</td><td colspan="3">Virtual interactive</td><td colspan="2">Physical</td></tr>
<tr><td>IS theory type</td><td>Analysis</td><td colspan="3">Explanation</td><td colspan="2">Prediction</td><td colspan="3">Explanation & prediction</td><td>Design & action</td></tr>
<tr><td rowspan="3">IS research method</td><td>Speculation/ comment</td><td colspan="3">Frameworks /conc. model</td><td colspan="2">Library research</td><td colspan="3">Literature analysis</td><td>Case study</td></tr>
<tr><td>Survey</td><td colspan="3">Field study</td><td colspan="2">Field experiment</td><td colspan="3">Laboratory experiment</td><td>Mathematical model</td></tr>
<tr><td colspan="2">Qualitative research</td><td colspan="3">Interview</td><td colspan="3">Secondary data</td><td colspan="2">Content analysis</td></tr>
</table>

4 Results

To summarize the insights gained from the literature review, we created a concept matrix (Table 4, characteristics not present in the literature database were omitted). In the following, we present the results of this structured literature analysis in detail.

Concerning the primary mode of communication, all three interaction types (text-based, speech-based, and combined) are addressed. Text-based CA (16 of 36) were for example explored in the context of user information disclosure behavior for sensitive topics [26], user perception of customer service agents with a focus on different agent- and communication-related cues [8], and the design of a natural search agent for legal research [29]. Research on CA that interact with users via speech (8 of 36) addressed for example lie detection by the CA [41] or service satisfaction with and continued use voice assistants [46]. Studies that focus on combined modes of interaction (12 of 36) include, for example, a study by Schroeder and Schroeder [11] comparing differences between interaction modes regarding users' willingness to share personal information and the development of overarching design principles for CA [4].

Table 4. Concept matrix

Article	CA mode			CA context		CA embodiment				IS theory type			IS research method							
	Text-based	*Speech-based*	*Both*	*General-purpose*	*Domain-specific*	*None*	*Virtual (static)*	*Virtual (interactive)*	*Physical*	*Analysis*	*Explanation*	*Design and Action*	*Case study*	*Interview*	*Laboratory experiment*	*Field experiment*	*Literature analysis*	*Secondary data*	*Speculation/commentary*	*Framework/conc. model*
[13]			X	X		X					X			X						
[47]	X			X		X					X				X					
[4]			X		X		X					X					X			
[48]			X	X		X				X								X		
[41]		X			X				X			X			X					
[49]			X		X		X					X			X					
[50]			X	X		X				X									X	
[51]			X	X		X					X				X					
[11]			X	X		X					X				X					
[12]	X			X		X					X				X					
[26]	X			X				X			X				X					
[25]			X	X				X			X				X					
[52]		X			X	X						X	X							
[29]	X				X	X						X	X							
[8]	X				X		X				X			X						
[53]	X				X	X						X			X					
[54]	X			X			X				X				X					
[55]	X				X	X						X			X					
[38]	X			X			X					X								X
[56]		X			X				X			X								X
[57]		X			X	X					X							X		
[58]			X	X		X					X						X			
[59]	X				X	X						X			X					
[60]			X		X	X					X							X		
[61]			X	X			X				X						X			
[46]		X		X		X					X									X
[62]		X			X			X				X		X						
[63]		X		X		X					X							X		
[64]		X			X	X					X				X					
[65]	X				X		X					X			X					
[66]	X				X	X					X				X					
[24]	X			X			X				X				X					
[40]			X		X		X				X				X					
[67]	X				X		X				X				X					
[21]	X				X	X					X					X				
[68]	X				X	X				X									X	
∑	**16**	**8**	**12**	**16**	**20**	**21**	**10**	**3**	**2**	**3**	**21**	**12**	**2**	**3**	**18**	**1**	**3**	**4**	**2**	**3**

With regard to the agent context, multiple studies research general-purpose CA (16 of 36) with a focus on aspects such as CA influence on decision-making [47] or the relation between CA capabilities and user experience [48]. Domain-specific CA (20 of 36) are studied in different contexts. Half of the studies on CA in specific domains focus on marketing and sales [21, 40, 49, 57, 59, 60, 64, 66–68], e.g. as product recommendation agents [40, 49] or in-store shopping assistants [21]. Furthermore, four studies investigate the use of CA in customer service [4, 8, 55, 56]. The six remaining domain-specific CA studies focus on a variety of specific contexts: Automated interviewing [41], gamified environments [52], workshop moderation [62], idea platforms [65], contracting [53], and legal research [29].

Concerning the embodiment of CA, most studies examine CA that do not have a virtual or physical representation (21 of 36), such as simple natural language systems [12, 29] or disembodied voice assistants [57, 64]. Further studies (10 of 36) explore CA with a virtual and static embodiment. For example, Wünderlich and Paluch [8] argue that the image of a CA represents an agent-related cue and describe its (potential) helpfulness for the user for perceiving a CA in a service encounter as authentic. Few studies address CA with a virtual interactive embodiment (3 of 36) or a physical embodiment (2 of 36). The three studies with a focus on CA with a virtual interactive representation indicate that interactive avatars can contribute to user affinity towards the CA through matching common human non-verbal cues [25], use facial expressions and gestures to increase perceived human-likeness [62], and might lead to more socially desirable responding for sensitive topics by users [26]. In addition, Nunamaker et al. [41] present a physically embodied CA with multiple sensors for interviewing and study different aspects of user interaction, such as perception of different CA genders or smiling. Finally, Stock and Merkle [56] study the use of a physically embodied, humanoid CA in a service encounter and find in their laboratory experiment that customer responded rather positively to innovative service behavior.

Regarding the produced IS theory types in the studies, most of the reviewed articles provide theories that intend to explain (21 of 36) different aspects of human-CA interaction, such as trust [44, 66], self-disclosure of information [66], or perceived authenticity [8]. Studies with a theoretical orientation towards design and action (12 of 36) for example provide design principles for CA in customer service [4] or for assistance in creative workshops [62], study the effect of dynamic response times on perceived social presence, perceived human-likeness and social presence [55] or evaluate anthropomorphic product recommendation agents [49]. Studies that focus on analysis (3 of 36) study capabilities of CA and user experience through the analysis of app reviews [48], consider the use of CA in e-commerce [68] or for information resource management [50].

The research methods used in the reviewed studies revealed a focus on laboratory experiments (18 of 36) which is one of most common methods in IS research [42, 43] and seems to be suitable to CA research as these experiments can provide useful insights into human-computer interaction. In addition, secondary data (4 of 36), such as data on consumption behavior for digital content through CA [57] or app reviews for virtual assistants [48] is used in the reviewed studies. Further research methods applied in the reviewed studies include the development of frameworks or conceptual models, interviews, literature analyses (each 3 of 36), case studies and speculation/commentary (each 2 of 36) as well as a field experiment by Al-Natour et al. [21].

5 Discussion

The goal of this literature review was to examine the status quo of CA within the field of IS research and provide directions for future work. In the following, we discuss our results and propose directions for prospective studies on CA.

5.1 State of CA Research in IS

Overall, our literature review shows a strongly increasing interest in CA in IS research, confirming similar statements by other scholars [4, 8, 55]. More than half of the identified studies were published in the last two years. The reviewed studies explore different communication modes, various application domains for CA and use different research methods to primarily produce theories for explanation and prediction as well as design and action.

Concerning the context in which the CA are studied, our results indicate a variety of different application domains of CA ranging from product recommendations [40, 49] over workshop moderation [62] to the use of CA for legal research [29]. In particular, marketing and sales (10 of 20 domain-specific studies) as well as customer service (4 of 20 domain-specific studies) were explored frequently, which are the two most popular application domains in an enterprise context for text-based CA [69, 70]. However, studies on CA as collaborators in team settings, i.e. machines as teammates [2], are limited to a single study on virtual workshop assistance [62].

With regard to the theory types, we find that only 12 of 36 studies produce theories to inform and guide the design of CA whereas successful design represents a major challenge in practice [16, 77, 78]. Interestingly, nearly all design-oriented studies focus on domain-specific CA, such as the design principles formulated by Gnewuch et al. [4] for CA in customer service or by Al-Natour et al. [21] for CA as shopping assistants.

We further observe a research focus on disembodied CA or CA with a virtual static embodiment (31 of 36) whereas studies on CA with virtual interactive embodiment or physical embodiment remain sparse. Many of the reviewed studies examine text-based CA, or chatbots, that are typically represented by means of a static avatar or lack any form of (virtual) embodiment [4]. A notable exception is the study by Stock and Merkle [56] who explore the use of the humanoid robot Pepper in a service encounter. Finally, our review of the applied research methods shows a focus on laboratory experiments (16 of 36) while the exploration of CA in the field, for example with the help of case studies (2 of 36) or field experiments (1 of 36) is limited.

5.2 Research Directions

Based on our assessment of the status quo, we propose four directions for future CA studies in IS research (Table 5): First, we suggest to extend the investigated application domains with regard to team settings in which a CA serves as a collaboration partner. While CA have the potential to support collaborative work [2], for example through the ad-hoc provision of information, we found only one study that addresses this context by developing design guidelines for CA as assistants for workshop moderation [62].

Second, we propose to conduct more and contrast design-oriented studies due to the rather limited number of studies that inform CA design and the fact that design remains a major challenge in practice [16, 71, 72]. As the contexts in which CA design is investigated become increasingly diverse, we believe it can be useful to contrast design knowledge across different contexts to identify common design principles as well as domain-specific aspects. For example, CA design for marketing and sales could place emphasis on increasing the persuasiveness of the agent [73] while a CA in customer service could benefit from being particularly empathetic in service encounters [20].

As a third research direction, we recommend to specifically study CA with a virtual interactive or physical embodiment as only five of the reviewed studies investigate embodied CA. Compared to disembodied CA or agents with a virtual static representation, embodied CA offer an increased variety of cues, such as facial expressions or gestures, that impact user interaction and provide additional options for design [30]. As technological approaches to create virtual interactive avatars improve [25] and physically embodied CA emerge, such as SoftBank's Pepper, we argue that studying their special features becomes increasingly relevant.

Table 5: Research directions and rationale

Direction	Description	Rationale
Investigate CA in team collaboration settings	Exploration of CA as innovative collaboration partners in team settings	CA in collaboration are only addressed by a single study despite their potential
Conduct more and contrast design-oriented CA studies	Formulation of design principles and design as well as evaluation of expository instantiations	Lack of design-oriented studies and successful CA design as a practical challenge
Explore virtually or physically embodied CA	Study of interaction with and design of embodied, both virtually and physically, CA	Existing focus on disembodied or virtually static embodied CA with limited social cues
Study CA introduction and use in the field	Expansion of CA research to the field, for example through case studies or field experiments	Half of the reviewed studies apply laboratory experiments in controlled settings

Finally, we propose to investigate real-life settings, for example through case studies or field experiments as half of the identified studies used laboratory experiments for their work. From our point of view, these experiments are useful for investigating different aspects in user and CA interactions in controlled settings. Nonetheless, we believe that future research can benefit from applying additional methods in the field [42, 43] regarding actual CA introduction and usage in enterprises in real-world settings. With these methods, results from CA studies can also provide stronger practical insights and thus ensure the relevance of our work for practitioners.

6 Concluding remarks

The aim of this research essay was to assess the status quo of CA in IS research and derive directions for future work. We conducted a systematic literature review and analyzed 36 studies with regard to five dimensions: CA context, CA communication mode, CA embodiment, IS theory type, and IS research method. Based on our findings we propose to move CA research in IS forward by the investigation of CA in collaborative settings, a stronger focus on design-oriented research, the exploration of embodied CA, and the study of CA in the field. While our study contributes to the understanding of the current state of CA research, it will require future updating and re-analysis as new studies emerge. Furthermore, we deliberately included only IS research in our study without incorporating work from other disciplines, such as human-computer interaction or computer science, to capture the status quo in IS research. We suggest that future work expands the view towards these disciplines when investigating design of or interaction with CA.

Overall, we believe that studying CA is a valuable research endeavor. In particular due the increasing capabilities of these agents and the variety of applications in private and professional life, they are an interesting, dynamic phenomenon to investigate in the context of digital transformation and can provide insights into new forms of human-computer interaction.

References

1. Welch, C.: Google just gave a stunning demo of Assistant making an actual phone call, https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018.
2. Seeber, I., Bittner, E.A.C., Briggs, R.O., Druckenmiller, D., de Vreede, G.-J., de Vreede, T., Maier, R., Oeste-Reiß, S., Randrup, N., Schwabe, G., Söllner, M.: Machines as Teammates: A Collaboration Research Agenda. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). pp. 420–429 (2018).
3. Verhagen, T., van Nes, J., Feldberg, F., van Dolen, W.: Virtual customer service agents: Using social presence and personalization to shape online service encounters. J. Comput. Commun. 19, 529–545 (2014).
4. Gnewuch, U., Morana, S., Maedche, A.: Towards Designing Cooperative and Social Conversational Agents for Customer Service. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–13 (2017).
5. Biro, M.M.: Recruiting By Chatbot: 4 Ways HR Tech Can Take Us Higher, https://www.forbes.com/sites/meghanbiro/2017/10/06/recruiting-by-chatbot-4-ways-hr-tech-will-take-us-higher/#71d2158870e9.
6. Oracle: Can Virtual Experiences Replace Reality? The future role for humans in delivering customer experience, (2016).
7. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM. 9, 36–45 (1966).
8. Wünderlich, N. V., Paluch, S.: A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–11 (2017).

9. Traumer, F., Oeste-Reiß, S., Leimeister, J.M.: Towards a Future Reallocation of Work between Humans and Machines – Taxonomy of Tasks and Interaction Types in the Context of Machine Learning. Proc. Int. Conf. Inf. Syst. 1–11 (2017).
10. Morana, S., Friemel, C., Gnewuch, U., Maedche, A., Pfeiffer, J.: Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz. Wirtschaftsinformatik Manag. 5, 42–51 (2017).
11. Schroeder, J., Schroeder, M.: Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS) (2018).
12. Schuetzler, R.M., Grimes, G.M., Giboney, J.S., Buckman, J.: Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–16 (2014).
13. Abul, M., Siddike, K., Spohrer, J., Demirkan, H., Kohda, Y.: People's Interactions with Cognitive Assistants for Enhanced Performances. Proc. Hawaii Int. Conf. Syst. Sci. 9, 1640–1648 (2018).
14. Webster, J., Watson, R.T.R.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. Manag. Inf. Syst. Q. 26, xiii–xxiii (2002).
15. McTear, M., Callejas, Z., Griol, D.: The conversational interface. (2016).
16. Luger, E., Sellen, A.: "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 5286–5297 (2016).
17. Lee, S.Y., Choi, J.: Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. Int. J. Hum. Comput. Stud. 103, 95–105 (2017).
18. Vogel-Meijer, K.: KLM: Making airline customer service soar with Messenger, https://www.facebook.com/business/success/klm-messenger.
19. NextIT: Helping a railroad service conduct business, http://nextit.com/case-studies/amtrak.
20. Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., Akkiraju, R.: Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2018).
21. Al-Natour, S., Benbasat, I., Cenfetelli, R.T.: The Role of Design Characteristics in Shaping Perceptions of Similarity: The Case of Online Shopping Assistants. J. Assoc. Inf. Syst. 7, 821–861 (2006).
22. Hanus, M.D., Fox, J.: Persuasive avatars: The effects of customizing a virtual salespersons appearance on brand liking and purchase intentions. Int. J. Hum. Comput. Stud. 84, 33–40 (2015).
23. Liao, Q.V., Hussain, M.M., Chandar, P., Davis, M., Crasso, M., Wang, D., Muller, M., Shami, N.S., Geyer, W.: All Work and no Play? Conversations with a Question-and-Answer Chatbot in the Wild. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2018).
24. Quynh, N., Sidorova, A.: Understanding User Interactions with a Chatbot: A Self-determination Theory Approach. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–5 (2018).
25. Seymour, M., Riemer, K., Kay, J.: Interactive Realistic Digital Avatars - Revisiting the Uncanny Valley. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). pp. 547–556 (2017).
26. Schuetzler, R.M., Giboney, J.S., Grimes, G.M., Nunamaker, J.F.: The Influence of Conversational Agents on Socially Desirable Responding. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). pp. 283–292 (2018).

27. Featherman, M., Thatcher, J., Wright, R.T., Zimmer, J.C.: Design Principles for Special Purpose, Embodied, Conversational Intelligence with Environmental Sensors (SPECIES) Agents. AIS Trans. Human-Computer Interact. 3, 1–25 (2011).
28. Lee, K.M., Jung, Y., Kim, J., Kim, S.R.: Are physically embodied social agents better than disembodied social agents?. Int. J. Hum. Comput. Stud. 64, 962–973 (2006).
29. Sugumaran, V., Davis, K.: A Natural Language-Based Multi-Agent System for Legal Research. In: Proceedings of the Americas Conference on Information Systems (AMCIS) (2001).
30. Sah, Y.J., Peng, W.: Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Comput. Human Behav. 45, 392–401 (2015).
31. Appel, J., Von Der Pütten, A., Krämer, N.C., Gratch, J.: Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. Adv. Human-Computer Interact. (2012).
32. Qiu, L., Benbasat, I.: A study of demographic embodiments of product recommendation agents in electronic commerce. Int. J. Hum. Comput. Stud. 68, 669–688 (2010).
33. Al-Natour, S., Benbasat, I., Cenfetelli, R.: The role of similarity in e-commerce interactions: The case of online shopping assistants. SIGHCI 2005 Proc. (2005).
34. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. Proc. Eur. Conf. Inf. Syst. 9, 2206–2217 (2009).
35. Bandara, W., Gorbacheva, E., Beekhuyzen, J., Furtmueller, E., Miskon, S.: Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support. Commun. Assoc. Inf. Syst. 34, 154–204 (2015).
36. Levy, Y., Ellis, T.J.: A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. Informing Sci. J. 9, (2006).
37. AIS: Senior Scholars' Basket of Journals, https://aisnet.org/page/SeniorScholarBasket.
38. Seeger, A.-M., Pfeiffer, J., Heinzl, A.: Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–17 (2018).
39. Gregor, S., Jones, D.: The Anatomy of a Design Theory. J. Assoc. Inf. Syst. 8, 312–334 (2007).
40. Wang, W., Benbasat, I.: Trust in and Adoption of Online Recommendation Agents. J. Assoc. Inf. Syst. 6, 72–101 (2005).
41. Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., Patton, M.W.: Embodied Conversational Agent-Based Kiosk for Automated Interviewing. J. Manag. Inf. Syst. 28, 17–48 (2011).
42. Palvia, P., Leary, D., Mao, E., Midha, V., Pinjani, P., Salam, A.F.: Research Methodologies in MIS: An Update. Commun. Assoc. Inf. Syst. 14, 526–542 (2004).
43. Palvia, P., Kakhki, M.D., Ghoshal, T., Uppala, V., Wang, W.: Methodological and Topic Trends in Information Systems Research: A Meta-Analysis of IS Journals. Commun. Assoc. Inf. Syst. 37, (2015).
44. Gregor, S.: The Nature of Theory in Information Systems. Manag. Inf. Syst. Q. 30, 611–642 (2006).
45. Arnott, D., Pervan, G.: Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. J. Assoc. Inf. Syst. 13, 923–949 (2012).

46. Otoo, B.A., Salam, A.F.: Mediating Effect of Intelligent Voice Assistant (IVA), User Experience and Effective Use on Service Quality and Service Satisfaction and Loyalty. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–9 (2018).
47. Elson, J.S., Derrick, D.C., Ligon, G.S.: Examining Trust and Reliance in Collaborations between Humans and Automated Agents. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS). pp. 430–439 (2018).
48. Nguyen, Q.N., Sidorova, A.: AI capabilities and user experiences: a comparative study of user reviews for assistant and non-assistant mobile apps. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–10 (2017).
49. Qiu, L., Benbasat, I.: Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. J. Manag. Inf. Syst. 25, 145–182 (2009).
50. Ryan, J., Snyder, C.: Intelligent Agents and Information Resource Management. In: Proceedings of the Americas Conference on Information Systems (AMCIS) (2004).
51. Saffarizadeh, K., Boodraj, M., Alashoor, T.M.: Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 0–12 (2017).
52. Silva-Coira, F., Pedreira, O., Cortiñas, A.: Intelligent Virtual Assistant for Gamified Environments. In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS) (2016).
53. Yoon, V., Rubenstein-montano, B., Wilson, T., Lowry, S., Yoon, V., Wilson, T., Lowry, S.: Natural Language Interface for a Multi Agent System. In: Proceedings of the Americas Conference on Information Systems (AMCIS) (2004).
54. Schuetzler, R.M., Grimes, G.M., Giboney, J.S.: An Investigation of Conversational Agent Relevance, Presence, and Engagement. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–10 (2018).
55. Gnewuch, U., Morana, S., Adam, M.T.P., Maedche, A.: Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In: Proceedings of the European Conference on Information Systems (ECIS) (2018).
56. Stock, R.M., Merkle, M.: Customer Responses to Robotic Innovative Behavior Cues During the Service Encounter. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–17 (2018).
57. Son, Y., Wonseok, O.: " Alexa, Buy Me a Movie !": How AI Speakers Reshape Digital Content Consumption and Preference. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–17 (2018).
58. Rzepka, C., Berger, B.: User Interaction with AI-enabled Systems : A Systematic Review of IS Research. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–16 (2018).
59. Maaß, W., Kowatsch, T., Janzen, S., Varshney, U.: A natural language technology-enhanced mobile sales assistant for in-store shopping situations. In: Proceedings of the European Conference on Information Systems (ECIS) (2011).
60. Baier, D., Rese, A., Röglinger, M., Baier, D., Rese, A., Röglinger, M.: Conversational User Interfaces for Online Shops? A Categorization of Use Cases. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–17 (2018).
61. Stieglitz, S., Brachten, F., Kissmer, T.: Defining Bots in an Enterprise Context. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–9 (2018).
62. Strohmann, T., Fischer, S., Siemon, D., Brachten, F., Lattemann, C., Robra-Bissantz, S., Stieglitz, S.: Virtual Moderation Assistance: Creating Design Guidelines for Virtual

Assistants Supporting Creative Workshops. Proc. Pacific Asia Conf. Inf. Syst. 3580–3594 (2018).

63. Zhao, L., Lu, X., Hu, Y.: A Proposed Theoretical Model of Discontinuous Usage of Voice-Activated Intelligent Personal Assistants (IPAs) A Proposed Theoretical Model of Discontinuous. In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS) (2018).
64. Kim, D., Park, K., Park, Y.: Alexa, Tell Me More: The Effect of Advertisements on Memory Accuracy from Smart Speakers. In: Proceedings of the Pacific Asia Conference on Information Systems (PACIS) (2018).
65. Tavanapour, N., Bittner, E.A.C.: Automated Facilitation for Idea Platforms: Design and Evaluation of a Chatbot Prototype. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–9 (2018).
66. Al-Natour, S., Benbasat, I., Cenfetelli, R.: The Antecedents of Customer Self-Disclosure to Online Virtual Advisors. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–18 (2009).
67. Wang, W., Benbasat, I.: Attributions of Trust in Decision Support Technologies: A Study of Recommendation Agents for E-Commerce. J. Manag. Inf. Syst. 24, 249–273 (2008).
68. Yu, B., Vahidov, R.: Agents and E-commerce: Beyond Automation. In: Proceedings of the Americas Conference on Information Systems (AMCIS). pp. 1–11 (2014).
69. Araujo, T.: Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Human Behav. 85, 183–189 (2018).
70. Vaccaro, K., Agarwalla, T., Shivakumar, S., Kumar, R.: Designing the Future of Personal Fashion Experiences Online. In: Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 1–11 (2018).
71. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Case study-Embodied virtual agents: An analysis on reasons for failure. J. Retail. Consum. Serv. 19, 605–612 (2012).
72. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Animated conversational agents and e-consumer productivity. Inf. Manag. 54, 545–559 (2017).
73. Adler, R.F., Iacobelli, F., Gutstein, Y.: Are you convinced? A Wizard of Oz study to test emotional vs. Rational persuasion strategies in dialogues. Comput. Human Behav. 57, 75–81 (2016).

The Potential of Augmented Reality for Improving Occupational First Aid

Jennifer Fromm[1], Milad Mirbabaie[1], and Stefan Stieglitz[1]

[1] University of Duisburg-Essen, Department of Computer Science and Applied Cognitive Science, Duisburg, Germany
{jennifer.fromm,milad.mirbabaie,stefan.stieglitz}@uni-due.de

Abstract. Accidents at work can happen at any time, but employees often do not feel prepared for such situations and therefore provide inadequate first aid. Using the Design Science approach, we developed a concept for an Augmented Reality (AR) application which uses voice assistance and visual AR overlays to support occupational first aiders with instructions in accident situations. The concept was implemented in the form of a video prototype and evaluated through interviews. The evaluation revealed that the application would enable employees without first aid training to apply first aid measures quickly and correctly; thus, reducing the probability of serious consequential injuries and costs in connection with occupational accidents.

Keywords: Augmented Reality, Digital Work, Occupational First Aid, Design Science, Design Thinking

1 Introduction

Digital technologies have not only become an integral part of our daily lives but also continuously change the way we work [1]. As a result of this development, digital transformation defined as "the use of digital technologies to enable major business improvement [2]" has emerged as an important topic for academics and practitioners alike. A major challenge companies currently face, is the development of new digital business models and the optimization of internal work processes through the use of emerging technologies [3]. These emerging technologies include, for example, Augmented Reality (AR) which allows virtual elements to be displayed in a person's field of vision [4]. Large companies like Facebook are already investing a lot in the further development of this technology, which is why AR has evolved into a trend topic [5]. According to the Gartner Hype Cycle, however, AR has already exceeded the peak of inflated expectations and it is therefore time to realistically evaluate the added value of the technology [6]. In this paper, we focus on the potential of AR instructions for improving the first aid measures provided after work-related accidents. We argue that using technology support in this area represents an opportunity for major business improvement, as the direct U.S. workers compensation costs of the ten most disabling workplace injuries in 2017 accounted for nearly 60 billion US dollars [7]. After an

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

accident at work, the immediate provision of appropriate first aid measures is of high importance for the prevention of more severe long-term injuries associated with higher costs [8]. Nevertheless, studies have repeatedly shown that appropriate first aid is often not provided because of a lack of first aid knowledge [9, 10].

We therefore used the Design Science approach [11] to develop a concept for solving this organizational problem. Early studies have already revealed that AR instructions can improve task performance and reduce mental workload while performing assembly tasks [12–14]. We propose that AR instructions could be well suited to guide workers while performing occupational first aid as well. The AR guidance could reduce uncertainties of workers while performing first aid leading to improved first aid measures and lower costs as a result of work-related accidents. The first goal of our research was to propose occupational first aid support as an additional use case for AR. In this regard, we developed a concept describing which functionalities an AR application would need to improve occupational first aid. The second goal of our research was to evaluate whether the proposed concept could solve the organizational problem of insufficient first aid measures after work-related accidents. Therefore, we implemented the app concept in the form of a video prototype and conducted evaluation interviews to receive feedback from potential users. Hence, we examine the following research question: **How can AR be used to improve occupational first aid?**

In the next chapter, we provide a theoretical background on occupational first aid and AR. Then, we describe how we developed the AR application concept and the corresponding video prototype by conducting a design thinking workshop. Afterwards, we present our concept and explain how we used the video prototype in combination with interviews to evaluate our concept. We conclude with a presentation of the evaluation results and a discussion on the added value of our AR concept.

2 Background

2.1 Occupational First Aid

In Germany, every person is legally obligated to provide first aid and non-assistance of an injured person can be sanctioned with imprisonment of up to one year or a fine (§323 StGB). Furthermore, employers are required to ensure that first aid is provided to their employees in case of an accident at work (§10 ArbSchG). Depending on the size and industry of a company, up to 10% of employees must complete a first aid course and repeat this course every two years (§26 DGUV Vorschrift 1). Similar legal regulations exist in other countries as well (e.g. US Occupational Safety and Health Act).

The training of occupational first aiders is essential because the survival rate and severity of an injury caused by an accident at work strongly depend on the provision of immediate and correct first aid measures. In this regard, it was already demonstrated that the provision of appropriate first aid measures can significantly reduce the hospital stay of workers who suffered from chemical burns after an accident at work [8]. The importance of performing cardiopulmonary resuscitation (CPR) correctly and immediately after a cardiac arrest to increase the injured person's survival chance has also been emphasized in prior research [15]. Occupational first aid training can contribute to preventing more severe injuries or casualties after an accident at work. Several studies revealed that bystanders who are trained in first aid respond more frequently in medical emergency situations and provided first aid measures of higher quality [16, 17]. However, other studies revealed that bystanders are often hesitating to provide first aid for several reasons such as fear of liability, lack of confidence, fear of making things worse and anxiety about performing in front of an audience [18]. Related to this, the usefulness of first aid training is often questioned as many studies show a quick decline of first aid knowledge and skills shortly after completing a first aid course. In a study with occupational first aiders, only 12% of the participants were able to perform CPR correctly six months after participating in a certified first aid course [9]. Similar results were also obtained in a more recent study in which only 5% of industry workers who completed a first aid course were able to answer questions about the estimation of vital signs and performance of CPR [10]. Moreover, a simulated scenario showed that trained participants tried to respond in a medical emergency situation but performed inappropriate first aid measures further damaging the injured person [19]. The researchers who conducted the aforementioned studies concluded that the repetition of a first aid course every two years is not sufficient to ensure an immediate and appropriate medical emergency response at work.

Summarized can be said that most research on occupational first aid comes from the medical field and focuses on assessing the effectiveness of first aid training and improving education in first aid. However, prior studies have continuously shown a quick decline of first aid knowledge and skills after course completion. Therefore, we assume that providing occupational first aiders with AR instructions during a medical emergency situation might be a more effective approach to ensure an immediate and appropriate response to an accident at work.

2.2 Augmented Reality

Milgram and Kishino [20] describe a virtual environment as “one in which the participant observer is totally immersed in and able to interact with, a completely synthetic world”. Information systems researchers agree that virtual environments enable users to experience a higher level of immersion, presence and interactivity than other systems [21]. Not all technologies, however, are able to present virtuality to the same degree. Thus, Milgram and Kishino [20] align physical reality, mixed reality and virtual reality along the so-called virtuality continuum. In physical reality, only objects that exist in the real world can be perceived, while virtual reality entirely consists of objects that do not exist in the real world. On the virtuality continuum, mixed reality is located in between and defined as an environment “in which real world and virtual world objects are presented together within a single display” [20]. Thus, AR becomes a subset of mixed reality which 1) combines real and virtual objects, 2) is interactive in real time and 3) registered in three dimensions [4]. This definition illustrates that the user of an AR application still maintains a sense of presence in the real world while the perception of his or her environment is enhanced by virtual objects [22]. Due to these characteristics, AR is most useful when “the success [of a task] is increased or made more likely (…) through additional visual information being presented alongside the physical world" [22].

A systematic literature review on AR in the field of information systems revealed that many previous studies focused on the use of the technology in the context of gaming, communication and collaboration processes, e-commerce, manufacturing, maintenance, education and health care [23]. Some of these studies have already demonstrated how the use of AR can lead to major business improvements. For example, Niemöller et al. [24] identified 38 work processes in the logistics industry which could benefit from the use of smart glasses. Furthermore, prototypical AR systems for providing manufacturing workers with assembly instructions have been implemented [13, 14]. An experimental study also revealed that AR instruction systems could improve task performance and reduce mental workload while performing assembly tasks [12]. In this paper, we propose that AR instructions could also be used to improve occupational first aid reducing consequential damage as a result of work-related accidents and costs for continued payment of compensation.

3 Research Approach

3.1 Design Science Research

The goal of Design Science Research (DSR) is the development and evaluation of an "IT artifact created to address an important organizational problem" [25]. According to March and Smith [26], artifacts can be either constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices) or instantiations (implementations and prototypes). In this research, we developed a concept for an AR first aid application and instantiated this concept in the form of a video prototype. In information systems research, prototypes have been defined as the "first embodiment of an idea" [27] and they should "capture the essential features of a later system" [28]. Most important is that an innovative concept can be tested and to accomplish this goal "neither the prototype's form nor the materials used in its construction have to be those of the final design" [29]. As the main goal of this research was not to explore the technological limitations of AR but to demonstrate an innovative AR use case and gain feedback from potential users, we considered the implementation of the AR application concept as a video prototype to be appropriate. During our research process, we adhered to the seven DSR guidelines by Hevner et al. [25] and followed the DSR methodology process model which is widely accepted in information systems research [11]. A key element of DSR is to use feedback gained through evaluation to further refine the initial design. Therefore, we plan to transform the video prototype into a functional prototype based on the evaluation results in future research. The methods we applied to perform each step of the DSR methodology process are summarized in Figure 1 and further explained in the following chapters.

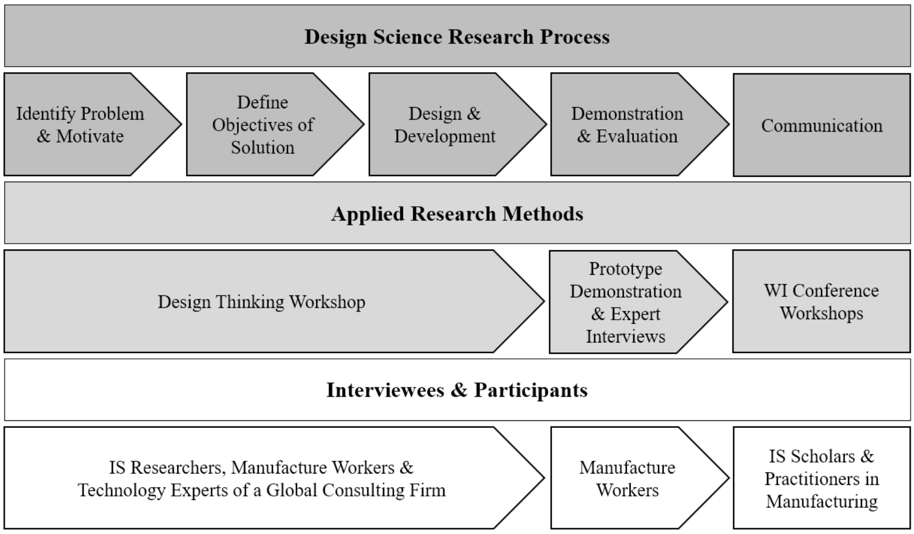


Figure 1. Research Design based on the DSRM Process Model [11]

3.2 Design Thinking Workshop

Though the DSR methodology process model specifies the different phases of DSR, "not much guidance is provided in information systems literature" [30] when it comes to how the goals of each phase can be achieved rigorously. Dolak, Uebernickel and Brenner [31], however, argue for an extension of DSR through Design Thinking (DT) and propose that considering DT as a valid DSR method would provide an added value for the field of information systems. We therefore decided to conduct a DT workshop with three teams including information systems researchers, manufacture workers and technology experts of a leading global consulting firm. We made the decision to invite manufacture workers as participants because the manufacturing sector belongs to the ones in which the highest number of accidents at work occur [32].

DT is a systematic, human-centered innovation approach utilizing design tools for solving organizational problems [33]. Brenner et al. [34] describe the DT process, methods and mindset as the three core elements of DT. A typical DT workshop is structured as follows [35]: The aim at the beginning of the DT process is to gain an understanding for a business problem and develop empathy for future users of the designed solution. Afterwards, ideas for a user-centered solution are developed and transformed into a prototype which is then evaluated with relevant stakeholders. Specific DT methods were developed for each phase of the DT process, for example, the empathy map technique is suitable for developing an understanding for stakeholders' needs in the initial phase [36]. The DT mindset frames the DT process and includes requirements for an effective collaboration such as multidisciplinary team constellations, unconstrained thinking and openness to unexpected ideas [37]. The structure of the DT workshop conducted in this research including the sequence of phases and applied methods can be found in Figure 2.

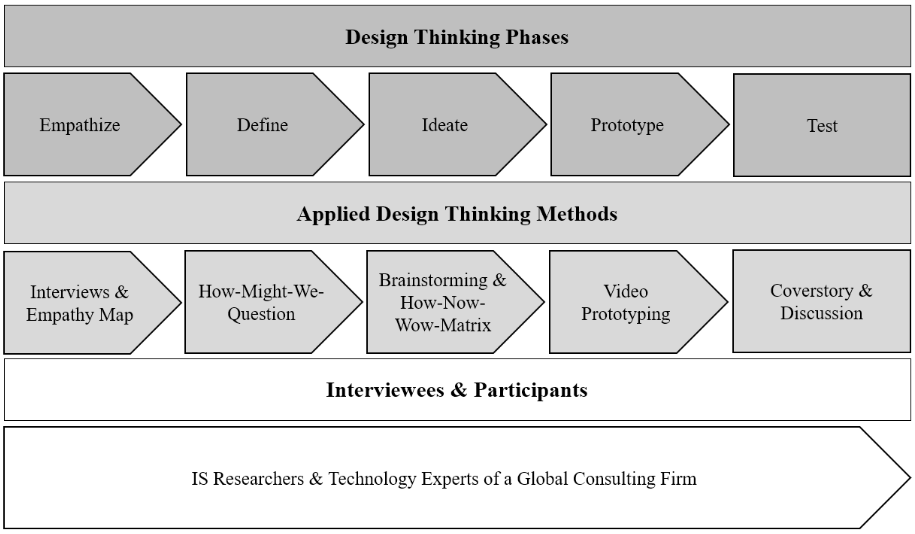


Figure 2. Structure of the Design Thinking Workshop Based on the Hasso-Plattner Institute D-SCHOOL Design Thinking Process [35]

The participants were informed in advance that an innovative AR solution for the improvement of occupational first aid measures should be developed during the workshop. In the first phase of the DT workshop, each researcher interviewed one of the participating manufacture workers using the empathy map technique. The workers were asked to share their experiences about accidents at work and the provided first aid. At the same time, the remaining team member took notes on the four quadrants of the empathy map: 1) what the person thinks and feels, 2) what the person sees, 3) what the person says and does and 4) what the person hears. For example, the first quadrant includes interviewee statements about thoughts and feels when he or she witnessed how a colleague had an accident at work.

The interviews revealed that work-related accidents occur frequently at the workplaces of the interviewees, although their employers invest a lot of money in occupational safety measures. Frequently mentioned work-related injuries were cuts, bruises and, in some cases, separated limbs because employees did not always wear protective clothing or disregarded road markings. Each participating worker had already experienced an accident at his or her workplace and was able to describe the treatment of an injured worker. In most cases, another person informed an occupational first aider who called an ambulance and provided first aid. The workers who participated in the DT workshop did not provide first aid in these situations because they were unsure whether they were allowed to do so without occupational first aid training and because they were not sure how they could help their injured colleague. Some interviewees also doubted that occupational first aiders are well prepared for every situation, as the obligatory first aid courses were described as rather short and superficial. All interviewees would welcome technological support in accident situations at the workplace. However, some thought that their employers would not be

willing to invest money in technological first aid support, as they already have to spend money on the obligatory training of occupational first aiders. Based on these insights, we defined the following how-might-we-question as guidance for the ideate phase:
How might we use AR to support occupational first aiders during an accident situation in a way that employers are willing to invest in this solution?

A brainstorming session followed in which each team generated ideas for an AR solution to address the problem stated in the how-might-we-question. To select one promising idea to pursue further, the teams made use of the how-now-wow-matrix. The x-axis of this matrix denotes the innovativeness of the ideas and the y-axis shows the feasibility. Ideas which were original but difficult to implement were sorted into the how-sector while the now-sector included ideas which were unoriginal but easy to implement. Most promising is the wow-sector including the most innovative and simultaneously feasible ideas. Each team then selected one idea of the wow-sector and created a video prototype demonstrating their idea to the other participants. As a last step, all teams imagined how their workplace could look like in the future after their solution has been implemented. Their visions were presented to the other participants in the form of a cover story about the successful implementation of their solution at their workplace. At the end of the workshop, the participants discussed which of the presented ideas provides the best solution for the problem stated in the how-might-we question and is technologically feasible as well. It was decided to further develop the idea of an AR mobile application which uses AR overlays and voice assistance to guide occupational first aiders in a medical emergency situation. The participants considered the combination of AR overlays and voice assistance as particularly innovative and assumed that this application would also be technologically feasible.

3.3 Further Development of the Artifact

After the workshop, the initial prototype of the AR mobile application was developed further. We used the design toolkit Sketch to design the user interface of the application. For each screen or interaction within a screen we created an artboard on a Sketch page. Then, we determined the logical user flow by linking artboards with their trigger action. We integrated pictures of a simulated accident situation into the Sketch artboards and used Photoshop to design AR overlays visualizing first aid instructions. Furthermore, we recorded ourselves to simulate the instructions of the voice assistant. The video prototype demonstrates how the application could be used to put an unconscious person in the recovery position. Of course, the final application should support employees in providing various other first aid measures. Screenshots of the visual first aid instructions demonstrated in the video prototype can be found in Figure 3. The figure also includes the voice assistant's instructions which are based on an online tutorial provided by the first aid charity organization St Johns Ambulance [38].

Figure 3. Screenshots of the Video Prototype Including Voice Assistant Instructions

3.4 Demonstration and Evaluation

To evaluate the usefulness of our artifact for the defined organizational problem, we demonstrated the video prototype to manufacture workers and conducted evaluation interviews. Though video prototypes do not yet provide complete functionality, different studies have shown that evaluation feedback gained through the demonstration of a video prototype does not significantly differ from the feedback gained through usability testing [39, 40]. Moreover, video prototypes have already been used in previous research to evaluate AR mobile applications [41, 42]. The interviews were transcribed and analyzed using the qualitative content analysis approach of Mayring [43] to assess the usefulness of the artifact and identify areas for further improvement. An overview about the interviewed manufacture workers can be found in Table 1.

Table 1. Description of Interviewees

No.	*Sex*	*Age*	*Job Title*
1	male	51	metal worker and safety coordinator
2	male	35	logistics specialist and safety coordinator
3	female	25	electronics technician
4	male	23	machine and equipment operator
5	male	25	car mechanic
6	male	25	assembly line worker

Prior to the interview, the participants were informed that they were about to evaluate an AR application for the assistance of occupational first aiders. Furthermore, we gave them a short introduction to AR and voice assistants. Then, the video prototype demonstrating the use of the application in an exemplary accident situation was shown to the participants. Afterwards, they were asked about their first impression of the presented application and to describe the functionalities of the application in their own words to identify potential misunderstandings. The interviewees then explained in detail what they like about the application and which improvements they would like to see. Furthermore, they indicated whether they would use the application.

4 Description of the AR First Aid Application Concept

The mobile AR application ImmediAID supports employees in providing first aid during medical emergency situations at the workplace. It is envisaged that an integrated voice assistant gives the user instructions and that these are visualized through AR overlays, so that even employees without any knowledge about first aid would be able to help an injured colleague at the workplace. The proposed functionalities of ImmediAID are explained in the following:

Emergency call: In the case of an accident, the employee uses the voice command "I am in an accident situation!" to activate the integrated voice assistant. The voice assistant then asks the employee whether an ambulance is required. If the employee answers "Yes" an emergency call is automatically initiated, the employee reports the accident and stays in the line until the emergency call center personnel ends the call. If the employee answers "No", the emergency call is not initiated and the voice assistant of ImmediAID continues to assist the employee.

Injury detection: The integrated voice assistant is able to detect the type of injury to guide the employee through the necessary first aid measures. If an emergency call was initiated, the voice assistant uses natural language processing and an interaction model to analyze the call between employee and emergency call center personnel in the background. If the emergency call was not initiated, the voice assistant asks the employee closed questions to assess the type of accident situation.

AR first aid support: Once the voice assistant has detected the type of injury, the appropriate first aid measure is looked up in a database which should be created in cooperation with medical personnel. Then, the voice assistant helps the employee to provide first aid by giving verbal instructions. Now, the employee aims with the smartphone camera at the injured person. AR overlays are shown on the smartphone display to visualize the verbal instructions of the voice assistant or the emergency call center personnel. ImmediAID is, for example, able to show the employee how to place an injured person in the recovery position by visualizing the correct angle and position for each body part with AR overlays. The accuracy of the AR overlays is ensured by using the OpenPose key point detection for identifying the body parts and axes of the injured person. An accuracy analysis of OpenPose demonstrated that the algorithm based on convolutional neural networks classified human postures in 95% of all images included in the test dataset [44].

AR first aid training: ImmediAID cannot only be used when an accident has already happened but also to prepare employees for a case of emergency. The app includes several AR first aid training tasks, for example, employees can practice how to place an injured person in the recovery position or how to perform CPR. Similar as in a real accident situation, the employee aims with the smartphone camera at a training partner or dummy. Then, the voice assistant guides the employee through the first aid exercise and each step is visualized on the smartphone display with AR overlays.

5 Evaluation Results

The evaluation revealed that the majority of the interviewees could imagine using an AR first aid app in an accident situation at the workplace. All interviewees stated that they would have felt able to carry out the instructions correctly as the instructions of the voice assistant and the displayed AR overlays were clear and easy to understand. It was also mentioned that the app would enable even staff members without first aid training to help. One interviewee liked the fact that the correct position of body parts was displayed in the form of AR overlays, as this makes it easier to follow the instructions. Related to this, one interviewee rated it positively that not only a video tutorial was shown, but AR overlays were displayed directly on the affected person. Most interviewees said that they would be glad to have such an app in an emergency, as it would give them a feeling of security. In this context, one interviewee could imagine that a well-chosen voice of the assistant could have a calming effect on first responders. A further interviewee found the voice input practical, so that one does not have to use one's hands to operate the smartphone when providing first aid. More than half of the interviewees were convinced that their employer would invest in AR support for occupational first-aiders. The interviewees described that their employers are already investing a lot of money in the occupational safety of their employees as work-related accidents are always associated with high costs such as rising contributions to the employers' liability insurance association. At the same time, costs are incurred for recruiting and training new employees if an injured person is absent for a longer period due to severe consequential damage as a result of a work-related accident. In addition, frequent occupational accidents are subject to increased controls by occupational safety authorities which are associated with costs and obligations.

Overall, the app idea was rated very positively, nevertheless, some of the interviewees also made suggestions for improvement. Most interviewees described the sound of the voice assistant as too mechanical and choppy. In this regard, the choice of a calming voice was recommended to reduce panic in an emergency. Another interviewee stated that he would be able to follow the instructions better if the voice assistant would speak more slowly. Due to his red-green visual impairment, one interviewee had difficulties to see the red AR overlays while another interviewee mentioned that the AR overlays were difficult to see in bright sunlight. It was therefore suggested that the app should automatically detect the brightness, background color, and color of the affected person's clothing to dynamically adjust the color of the AR overlays. As the video prototype demonstrated how the app could support occupational first aiders in placing an unconscious person in the recovery position, several interviewees noted that the app should be able to support more typical workplace injuries. Specifically, instructions for the treatment of cuts, bruises, bone fractures, strains, foreign bodies in the eye, electric shocks and severed limbs were mentioned. It should be noted that the video was only a prototypical realization of the app concept and that the concept indeed envisaged the support of such typical injuries. Another interviewee wished for a simple possibility to repeat the last instruction by voice command (e.g. "Again!"). Of course, there might be situations in which the voice assistant may not be able to detect an injury or suggest a suitable first aid measure. In this case, it was important to an interviewee that the app would still give general instructions such as "Keep calm!" or "Do not leave the affected person alone!" instead of an error message. One interviewee explained that mobile phones are forbidden at her workplace and that there is no fast internet connection available. The interviewee suggested placing a charged mobile phone with the app next to first aid boxes or equipping employees with company mobiles. If the app would not be able to process the images quickly enough due to a slow internet connection, the interviewee further suggested that the app should preload instructions in the form of videos. An overview about the mentioned strengths of the presented AR first aid application and suggestions for improvement can be found in Table 2.

Table 2. Strengths and Improvement Suggestions

Main Category	*Sub Category*	*Occurrence*
Strengths	Clarity of instructions	6
Strengths	Reduction of uncertainties	4
Strengths	Visualization of the body parts' target position	2
Strengths	Visualization of the actually affected person	1
Strengths	Calming effect of the voice assistant	1
Strengths	Mobility through speech input	1
Improvement	Sound of the voice assistant	5
Improvement	Visibility of the AR overlays	2
Improvement	Instructions for other accident situations	2
Improvement	Voice command to repeat the last instruction	1
Improvement	App behavior in case of malfunctions	1
Improvement	Necessity of mobile phones and internet	1

6 Conclusion

The aim of this research was to propose and evaluate occupational first aid support as an innovative use case of AR. We therefore conducted a DT workshop with information systems researchers, manufacture workers and technology experts of a global consulting firm to develop a concept for an AR mobile first aid application. The proposed application supports occupational first aiders by providing instructions through visual AR overlays and voice assistance. We further proposed to use the OpenPose key point detection algorithm to locate body parts of an injured person and calculate the optimal position for AR overlays. The concept was implemented as a video prototype and evaluated by conducting interviews with manufacture workers. The evaluation revealed that AR instructions would enable employees without first aid training to act quickly and correctly in accident situations. The interviewees were also convinced that their employer would invest in such a solution because inadequate first aid measures are often associated with long-term staff absenteeism and thus incur high costs.

The evaluation interviews further revealed some challenges which have to be addressed before the proposed application can be used in practice. The most critical challenge will be to ensure accessibility to the application in medical emergency situations. An interviewee reported that the use of smartphones is forbidden at her workplace and that it would be difficult to provide first aid measures while pointing the smartphone camera at an injured person. A solution would be to use smart glasses, as particularly manufacturing and logistics companies are already using these to improve the efficiency of internal work processes. The proposed application should further be able to identify the required first aid measure by analyzing the camera recordings and voice inputs of the user. To ensure a high accuracy, medical personnel should be involved in the development process of the application. Nevertheless, innovative ideas involving artificial intelligence always raise questions of liability which must be addressed in future research. Who is liable if an injured person dies or takes further damage due to false recommendations of an artificial intelligence? Related to this question, the interviewees stated that in most accident situations they would not be able to provide the correct first aid measure on their own – so they would be willing to use the application even if following the instructions might be harmful in some cases.

The AR application concept was developed as part of a DT workshop with practitioners. DT is a popular innovation approach that promotes creative thinking, however, the approach prescribes setting fixed time limits for each DT method. As a result, ideas are often not yet entirely mature after a workshop. Although we have further refined the concept and the video prototype after the workshop, the implementation of the concept as a video prototype represents a limitation. We therefore plan to implement a functional prototype in future research to evaluate the technological possibilities and limitations of AR for occupational first aid support. The implementation of a functional prototype would allow to assess the accuracy of the OpenPose key point detection, position of the AR overlays and recognition of required first aid measures. The functional prototype should also be used to quantitatively measure whether first aid instructions are implemented faster and more correctly than without the application. Moreover, it should be tested how different representations of AR overlays affect the correctness and timeliness of provided first aid measures. In this regard, the evaluation interviews already revealed that the brightness and coloring of visual AR overlays plays an important role and should adapt dynamically.

References

1. Colbert, A., Yee, N., George, G.: The Digital Workforce and the Workplace of the Future. Acad. Manag. J. 59, 731–739 (2016)
2. Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M.: Embracing Digital Technology: A New Strategic Imperative. MIT Sloan Manag. Rev. 55, 1–14 (2014)
3. Hanelt, A., Piccinini, E., Gregory, R.W., Hildebrandt, B., Kolbe, L.M.: Digital Transformation of Primarily Physical Industries - Exploring the Impact of Digital Trends on Business Models of Automobile Manufacturers. In: 12th International Conference on Wirtschaftsinformatik, p. 88. AIS, Osnabrück (2015)
4. Azuma, R.T.: A Survey of Augmented Reality. Teleoperators Virtual Environ. 6, 355–385 (1997)
5. Facebook, https://de.newsroom.fb.com/news/2018/05/f8-2018-frei-zugaengliche-ki-frameworks-neue-entwicklungen-in-ar-vr-und-weitere-highlights-des-zweiten-tages/ (Accessed: 12.11.2018)
6. Gartner, https://www.gartner.com/doc/3885468/hype-cycle-emerging-technologies (Accessed: 12.11.2018)
7. Liberty Mutual Insurance, https://business.libertymutualgroup.com/business-insurance/Documents/Services/Workplace%20Safety%20Index.pdf (Accessed: 12.11.2018)
8. Leonard, L.G., Scheulen, J.J., Munster, A.M.: Chemical Burns: Effect of Prompt First Aid. J. Trauma. 22, 420–423 (1982)
9. McKenna, S.P., Glendon, A.L.: Occupational First Aid Training: Decay in Cardiopulmonary Resuscitation (CPR) Skills. J. Occup. Organ. Psychol. 58, 109–117 (1985)
10. Hatzakis, K.D., Kritsotakis, E.I., Angelaki, H.P., Tzanoudaki, I.K., Androulaki, Z.D.: First Aid Knowledge Among Industry Workers in Greece. Ind. Heal. 43, 327–332 (2005)
11. Pfeffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 24, 45–77 (2007)
12. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative Effectiveness of Augmented Reality in Object Assembly. In: 21st Conference on Human Factors in Computing Systems. pp. 73–80. ACM, Fort Lauderdale (2003)
13. Caudell, T.P., Mizell, D.W.: Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes. In: 25th Hawaii International Conference on System Sciences. pp. 659–669. IEEE, Kauai (1992)
14. Saggiomo, M., Loehrer, M., Kerpen, D., Lemm, J., Gloy, Y.: Human- and Task-Centered Assistance Systems in Production Processes of the Textile Industry: Determination of Operator-Critical Weaving Machine Components for AR-Prototype Development. In: 49th Hawaii International Conference on System Sciences. pp. 560–568. IEEE, Koloa (2016)
15. Gilmore, C.M., Rae, T.D., Becker, L.J., Eisenber, M.S.: Three-Phase Model of Cardiac Arrest: Time-Dependent Benefit of Bystander Cardiopulmonary Resuscitation. Amer. J. Cardiol. 98, 497–499 (2006)
16. Bakke, H.K., Steinvik, T., Eidissen, S.I., Gilbert, M., Wisborg, T.: Bystander First Aid in Trauma - Prevalence and Quality: A Prospective Observational Study. Acta. Anaesthesiol. Scand. 59, 1187–1193 (2015)

17. Tanigawa, K., Iwami, T., Nishiyama, C., Nonogi, H., Kawamura, T.: Are Trained Individuals More Likely to Perform Bystander CPR. Resuscitation. 82, 523–528 (2011)
18. Das, M., Elzubeir, M.: First Aid and Basic Life Support Skills Training Early in the Medical Curriculum: Curriculum Issues, Outcomes, and Confidence of Students. Teach. Learn. Med. 13, 240–246 (2001)
19. Bakke, H.K., Steinvik, T., Angell, J., Wisborg, T.: A Nationwide Survey of First Aid Training and Encounters in Norway. BMC Emerg. Med. 17, 1–7 (2016)
20. Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. Trans. Inf. Sys. 77, 1321–1329 (1994)
21. Walsh, K.R., Pawlowski, S.D.: Virtual Reality: A Technology in Need of IS Research. Commun. Assoc. Inf. Syst. 8, 20 (2002)
22. Steffen, J.H., Gaskin, J., Meservy, T., Jenkins, J.: The Missing Framework for Virtually Assisted Activities. In: 38th International Conference on Information Systems. AIS, Seoul (2017)
23. Harborth, D.: Augmented Reality in Information Systems Research: A Systematic Literature Review. In: 23rd Americas Conference on Information Systems. AIS, Boston (2017)
24. Niemöller, C., Zobel, B., Berkemeier, L., Metzger, D., Werning, S., Adelmeyer, T., Ickerott, I., Thomas, O.: Sind Smart Glasses die Zukunft der Digitalisierung von Arbeitsprozessen? Explorative Fallstudien zukünftiger Einsatzszenarien in der Logistik. In: 13th International Conference on Wirtschaftsinformatik. pp. 410–424. AIS, St. Gallen (2017)
25. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science Research in Information Systems Research. MIS Q. 28, 75–105 (2004)
26. March, S.T., Smith, G.F.: Design and Natural Science Research on Information Technology. Decis. Support Syst. 15, 251–266 (1995)
27. Glegg, G.L.: The Development of Design. Cambridge University Press, Cambridge (1981)
28. Naumann, J.D., Jenkins, A.M.: Prototyping: The New Paradigm for Systems Development. MIS Q. 6, 29–44 (1982)
29. Janson, M., Smith, L.: Prototyping for Systems Development: A Critical Appraisal. MIS Q. 9, 305–316 (2008)
30. Offermann, P., Levina, O., Schönherr, M., Bub, U.: Outline of a Design Science Research Process. In: 4th International Conference on Design Science Research in Information Systems and Technology. ACM, Philadelphia (2009)
31. Dolak, F., Uebernickel, F., Brenner, W.: Design Thinking and Design Science Research. In: 8th International Conference on Design Science Research in Information Systems and Technology. ACM, Helsinki (2013)
32. European Commission, https://ec.europa.eu/eurostat/documents/3859598/5926181/KS-RA-12-102-EN.PDF/56cd35ba-1e8a-4af3-9f9a-b3c47611ff1c (Accessed: 12.11.2018)
33. Brown, T.: Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation. HarperCollins, New York (2009)
34. Brenner, W., Uebernickel, F., Abrell, T.: Design Thinking as Mindset, Process, and Toolbox. In: Brenner, W., Uebernickel, F. (eds.) Design Thinking for Innovation. pp. 3–21. Springer (2016)
35. HPI-D-SCHOOL, https://hpi.de/school-of-design-thinking/design-thinking/hintergrund/design-thinking-prozess (Accessed: 12.11.2018)
36. Tschimmel, K.: Design Thinking as an Effective Toolkit for Innovation. In: 23th International Society for Professional Innovation Management Conference. ISPIM, Barcelona (2012)
37. Carlgren, L., Rauth, I., Elmquist, M.: Framing Design Thinking: The Concept in Idea and Enactment. Creat. Innov. Manag. 25, 38–57 (2016)

38. St John Ambulance, http://www.sja.org.uk/sja/first-aid-advice/first-aid-techniques/the-recovery-position.aspx (Accessed: 12.11.2018)
39. Zwinderman, M., Leenheer, R., Shirzad, A., Chupriyanov, N., Veugen, G., Zhang, B., Markopoulos, P.: Using Video Prototypes for Evaluating Design Concepts with Users: A Comparison to Usability Testing. In: 15th International Conference on Human-Computer Interaction. pp. 774–781. ACM, Las Vegas (2013)
40. Dhillon, B., Banach, P., Kocielnik, R., Emparanza, J.P., Politis, I., Pkaczewska, A., Markopoulos, P.: Visual Fidelity of Video Prototypes and User Feedback: A Case Study. In: 14th International Conference on Human-Computer Interaction. pp. 139–144. ACM, Orlando (2011)
41. Rane, P., Kim, H., Marcano, J.L., Gabbard, J.L.: Virtual Road Signs: Augmented Reality Driving Aid for Novice Drivers. In: 60th Human Factors and Ergonomics Society Annual Meeting. pp. 1750–1754. HFES, Washington (2016)
42. de Sá, M., Antin, J., Shamma, D.A., Churchill, E.: Mobile Augmented Reality: Video Prototyping. In: 29th Conference on Human Factors in Computing Systems. pp. 1897–1902. ACM, Vancouver (2011)
43. Mayring, P.: Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Beltz, Klagenfurt (2014)
44. Ghazal, S., Khan, U.S.: Human Posture Classification Using Skeleton Information. In: 1st International Conference on Computing, Mathematics and Engineering Technologies. pp. 1–4. IEEE, Sukkur (2018)

Prevent a Vicious Circle!

The Role of Organizational IT-Capability in Attracting IT-affine Applicants

Andrea Fürst-Graßl[1]

[1] University of Passau, Chair of Business Information Systems, Passau, Germany
andrea.fuerst@uni-passau.de

Abstract. Organizational IT-capability is a key to meet the challenges of digital transformation. Its advancement requires the attraction of IT-affine applicants. Their application intentions, however, depend on positive perceptions of an employer's existing IT-capability. Merging findings from Information Systems Research, Human Resource Research, and Psychology, this study conducts an experimental survey to investigate how the perception of organizational IT-capability influences job seekers' application intentions. 228 observations were analyzed using the PLS-SEM method. It provides empirical evidence that (1) the perception of organizational IT-capability influences job seekers' application intentions and that (2) this influence is of particular strength for IT-affine job seekers. This knowledge is fundamental to understand the outstanding role of organizational IT-capability in the current 'war for talents'.

Keywords: Digital Transformation, IT-Capability, Person-Organization Fit, Intention to Apply, Applicants' Self-Selection

1 Introduction

Organizational Information Technology (IT)-capability is considered a key factor to transform IT into a competitive advantage [1–3]. In a Resource Based View, organizational IT-capability is a complex package of IT-related resources including human skills and knowledge [4, 5]. As employees' knowledge and skills play a major part in the enhancement of a firm's IT-capability it is crucial for firms to attract and employ technically high-skilled employees [4].

Current shortages on the labor market, however, are creating a 'war for talents', especially with regard to high-skilled IT-specialists. A managerial study of Capgemini Consulting, for example, revealed that "77% of [the] companies considered missing digital skills as the key hurdle to their digital transformation" [6]. Therefore, it is crucial for organizations to know about the specific drivers of IT-affine job seekers' application intentions.

Following P-O-fit theory, job seekers generally prefer organizations that match their individual characteristics [7–11]. This fosters a feeling of 'fit' with a potential

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

employer, which was found to be essential for application intentions [10]. Applying this fit-perspective, it seems especially important for IT-affine job seekers that their individual IT-capability matches the potential employer's organizational IT-capability. Consequently, job seekers with a high level of individual IT-capability are likely to prefer organizations with an equally high level of organizational IT-capability, whereas they are likely to eschew employers with a minor level of IT-capability.

This endangers low-capability organizations to enter a vicious circle: From a resource based perspective, the development of their organizational IT-capability requires the attraction of high-skilled employees [4] – which they probably cannot attract. Consequently, their organizational IT-capability will further diminish, which makes it more and more difficult to attract savvy employees. To prevent such a vicious circle, it is crucial to empirically investigate the underlying mechanisms to develop proper countermeasures.

Therefore, this study addresses two primary research questions:

1. Does the perception of organizational IT-capability influence job seekers' intention to apply?
2. Is this influence of particular strength for job seekers with a high level of individual IT-capability?

To answer these questions this research takes a new perspective on IT-Capability and combines it with existing theories and metatheories anchored in Human Resource Research and Psychology, (1) to investigate how the perception of organizational IT-Capability influences job seekers' application intention and (2) to explain applicants' self-selection behavior in this context. From a practical point of view, this knowledge is fundamental to understand the outstanding role of organizational IT-capability in the current 'war for talents'.

2 Theoretical Background and Hypotheses Development

2.1 Theoretical Framing

Previous research pointed out that the attraction of potential applicants is – from a theoretical point of view – a complex issue that results from an interaction of various theories concerning information processing, perception development and subjective fit [12]. Therefore, this study does not built on a single theory but merges findings from different relevant theories.

This merge primarily bases on the directive theoretical contribution of Ehrhart and Ziegert [12] who presented relevant theories from human resource research and psychology and organized them along their key-aspects in respective meta-theories. Especially the first two meta-theories provide a valuable theoretical background for this study: The first meta-theory – *Environmental-Processing Theory* (including, inter alia, *image theory* and *signaling theory*) emphasizes that (1) there is a difference between objective and individually perceived characteristics of an organization and that (2) there is a relationship between those perceptions of organizational characteristics and attraction [12]. The second meta-theory – *Interactionist Processing Metatheory* – strongly builds on *P-O-fit theory* [9] and primarily considers the interaction between person characteristics and environmental characteristics [12]. Both meta-theories (and their underlying theories) feed in the subsequent hypotheses development sections.

Additionally, research from the field of value-driven marketing (see [13] for an overview) is taken into account to further support the basic idea of this study that particularly *perceptions* of organizational IT-Capability matter for the attraction of proper applicants – as this research stream clearly points out, that the value of a product or service process (including a vendor's resources and capabilities) depends on the customer's perception and appreciation [14].

2.2 Organizational IT-Capability and Intention to Apply

Organizational IT-capability "is a firm's ability to acquire, deploy, combine, and reconfigure IT-resources in support and enhancement of business strategies and work processes" [2]. Following the established conceptualization of IT-capability in Information Systems Research, it reflects the three dimensions: IT Infrastructure Capability, IT Business Spanning Capability, and IT Proactive Stance [2]. IT Infrastructure Capability captures the extent to which a firm is good at data management, network communication, architecture and application management, IT Business Spanning Capability considers a successful interaction between business and IT strategy, and IT Proactive Stance reflects, whether a firm is open for IT-innovations [2].

As an organization's IT-capability reflects a complex interaction between these three dimensions including technical and human resources, it is rare and difficult to imitate [4]. It was found to be an important enabler of organizational agility [2] and a key factor to meet the current challenges of digital transformation [15]. Therefore, previous research pointed out, that organizational IT-capability constitutes a competitive advantage [4]. As it determines an employer's competitive standing and sustainability it may be of high relevance for potential employees.

An organization's IT-capability is, however, not an externally observable attribute. Outsiders can only develop perceptions of an organization's IT-capability based on the provided information. This information can, for example, be included into job ads – which were identified as an important source of information for applicants [16].

However, as an organization's IT-capability must be kept rare and inimitable to further constitute a competitive advantage, an organization can not disclose any IT-related information. It will probably not disclose detailed insider knowledge about its database technology, application landscape, IT-architecture, or future IT-related business options. It can, however, provide rather general IT-related information that reflects a certain level of IT-expertise without jeopardizing its competitive advantage. For example, an organization could provide general information about the extent to which IT is integrated in its products, services, and business processes, whether they align business and IT strategy or about its fundamental stance towards IT innovations.

Following Signaling Theory, this information serves as a signal for potential applicants to develop perceptions of the respective organization's IT-Capability. [12, 17]. This *Perceived Organizational IT-Capability* can be defined as the degree to which a person believes that an organization is capable of using IT to support its business strategy and processes. This perceived organizational IT-capability may influence job seekers' application intentions for two main reasons:

1. Following Social Identity Theory, individuals prefer organizations they identify as superior to appear more successful in the eyes of others [10, 18]. Following Image Theory individuals prefer organizations that may foster a desired image [19, 20]. Building on these theories, previous studies for example revealed a clear relationship between a superior employer reputation [21–25] or a superior corporate image [10, 26, 27] and job seekers' application intentions. Therefore, a (perceived) superior level of organizational IT-capability is likely to positively influence job seekers application intentions as well.
2. As an organization's IT-capability constitutes a competitive advantage and is the key to manage the current challenges of digital transformation [2, 4, 15], it determines an employer's long-term sustainability and future economic success. It, therefore, ensures security of employment and employee salaries and provides the base for advancement possibilities. As these are important issues for potential employees [7, 10, 16], it is likely that they prefer employers they perceive as high-capable for this reasons as well.

Therefore I hypothesize:
H1: Perceived Organizational IT-Capability is positively associated with job seekers' Intention to Apply.

2.3 The Moderating Role of Individual IT-capability

The term IT-capability is usually used in an organizational context. However, the overall ability to acquire and purposefully use IT is not limited to organizations. In fact, each job seeker is characterized by a certain level of individual IT-capability. Reflecting his IT affinity in a professional sense, this *Individual IT-Capability* can be defined as a job seeker's ability to professionally use IT to support an employer's business goals.

This individual characteristic may affect the strength of the hypothesized influence of *Perceived Organizational IT-Capability* on *Intention to Apply*. This influence is supposed to be of particular strength for job seekers with a high level of *Individual IT-Capability* and weaker for job seekers with a rather low level of *Individual IT-Capability* for two main reasons:

1. Individual perception development generally depends on individual characteristics and preferences. People process information selectively and limit their attention to decision relevant aspects [12, 17, 19, 20]. Knowledgeable individuals were found to include more attributes in their perception development processes than less savvy individuals [28]. Therefore, IT-affine job seekers are likely to pay particular

attention to IT-related information about a potential employer and develop a sounder perception of a potential employer's organizational IT-capability than less savvy job seekers. This sounder perception of organizational IT-capability is likely to constitute a clearer influence of *Perceived Organizational IT-Capability* on *Intention to Apply*.
2. Person-Organization (P-O) Fit Theory emphasizes the importance of perceived similarity between a person and its working environment. Reflecting the perceived congruence between individual and organizational attributes, interests, and values, P-O fit turned out to significantly influence job seekers' application intentions and subsequent job choice decisions [8, 10, 29–31]. Consequently, job seekers with a high level of *Individual IT-Capability* are likely to clearly prefer employers with a similar high level of IT-capability which match their interests, use their IT-related skills and knowledge, and provide proper advancement opportunities. On the other hand, these IT-affine job seekers are likely to clearly eschew employers with an inferior level of IT-capability due to the missing 'fit'. In contrast, job seekers with a rather low level of *Individual IT-Capability* may be somewhat discouraged by an exceedingly high level of employer IT-capability, so that the hypothesized positive relationship between *Perceived Organizational IT-Capability* and *Intention to Apply* may be weaker for less IT-affine job seekers.

Therefore, I hypothesize:
H2: The relationship between Perceived Organizational IT-Capability and Intention to Apply is stronger for job seekers with a high level of Individual IT-Capability compared to job seekers with a minor Individual IT-Capability.

3 Research Method

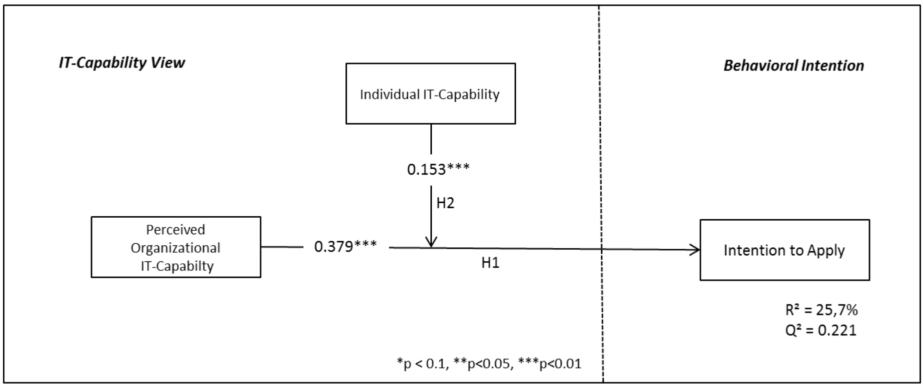


Figure 1. Research Model (Including Results of the PLS Estimation)

3.1 Study Design and Data Collection

To empirically test the hypothesized relationships this study applied an experimental survey approach. It was based on four hypothetical job ads: Whereas the company descriptions differed reflecting different levels of organizational IT-capability, all other factors (job description, company size and starting salary) were kept constant (see Appendix).

Confronting each participant with multiple hypothetical job ads seemed especially important as perception development processes were found to be comparative by nature [28]. Therefore, it seems – on the one hand – reasonable to confront each participant with multiple descriptions as a reference group to gain a proper assessment of each firm's IT-Capability. On the other hand, this research design may include the risk of gaining not entirely independent observations. To take both aspects into account I decided to confront each participant with four job ads – as a good middle – and to apply an additional test to ensure the validity of my results (see section 4.2).

Data were collected using two surveys (a pretest and a main survey) of students from the Department of Business, Economics and Information Systems of a German university during summer term 2017. To ensure a high response rate, the two surveys were conducted during class time in two IT-related classes. Business students of IT-related classes seemed a proper sample for three main reasons: First, they presumably have enough IT-related knowledge to assess their individual IT-Capability properly. Second, they are presumably more open-minded towards different job offers than professionals, as they are less influenced by prior work experiences. And third, this sample provides realistic and valuable results for the current labor as the questioned students are available for the labor market in the near future.

To ensure this last point, all participants were asked whether they can imagine applying for a job within the next twelve months at the beginning of the questionnaire. Only questionnaires with a positive answer to this question were taken into account and included in the further analysis. The participants were also questioned about their degree program, current semester, age and gender.

Thereupon, each participant was asked about a self-evaluation of his individual IT-Capability. This variable forms the base for the following moderation analysis. Afterwards, each participant was confronted with the aforementioned four hypothetical job ads reflecting different levels of organizational IT-capability. As previous research pointed out, that the sole quantity of information in job ads may influence job seekers' perceptions [16], particular attention was paid to a similar length of all provided company descriptions. These descriptions were all placed above the related questions to ensure that each participant has all descriptions in mind when answering the questionnaire. Each participant was then asked about his perception of each organization's IT-capability and his intention to apply to the respective organization.

The questionnaire (including the hypothetical job ads) was pretested for understandability and proper construct measurement with 16 students from a software development class. The pretest showed a good understandability of the provided information and the measurement scales as well.

The main survey was conducted in an information system class at the end of summer term 2017. From the main survey, I received 57 completed and usable questionnaires, each including assessments of the four different firms. This resulted in 57*4=228 total observations for further analysis. Among the 57 participants, 59.6% were female, 40.4% were male. The average age was 22.7 years. The participants were distributed over the Bachelor Degree Programs “International Cultural and Business Studies” (14.0%), “Business Administration and Economics” (78.9%) and “Information Systems” (7.0%).

As the study was conducted with German-speaking participants the questionnaire was provided in German language as well to prevent language-based problems. (The English translation of the provided information can be found in the Appendix). To ensure content equivalence of the German and English version, I followed the translation and back-translation procedure recommended in Brislin [32].

3.2 Construct Measurement

A job seeker’s ‘Intention to Apply’ (INTA) is an established and widely applied construct in recruitment research. Therefore, its measurement items were derived and adapted from the established measurement scales [10, 24, 33].

The established measurement scales for IT-Capability, however, are not entirely suitable for this study. The established view on IT-Capability is limited to organizations and demands deep insider knowledge about the respective firm. This study, however, requires – on the one hand – a measure to capture an outsider’s perception of an (potential) employer’s IT-Capability and – on the other hand – a measure that reflects each job seeker’s individual IT-Capability. These ‘views’ on IT-Capability both differ from the established conceptualization of IT-Capability and its related measurement scales. Therefore, this study develops two ‘new’ constructs *Perceived Organizational IT-Capability* (POITC) and *Individual IT-Capability* (IITC) – both derived from the ‘traditional’ IT-Capability perspective. This development procedure was realized in several steps:

1. To form a theoretical foundation for the formulation of proper measurement items, relevant prior research was reviewed – particularly focusing on different conceptualizations of IT-Capability, perception development processes, and individual aspects (e.g., the individual attitude towards IT). Thereupon, reflective measurement scales were formulated, which seemed suitable to capture a job seeker’s perception of an organization’s IT-Capability and to give an analogous self-evaluation of his individual IT-Capability.
2. These self-developed measures were afterwards discussed with several experienced researchers to ensure a proper measurement design with regard to explanatory value and understandable formulation. Adaptions were made where necessary.
3. Finally, a pretest was conducted with 16 students, to further evaluate the developed indicators’ understandability, validity and reliability. After finishing the pretest-questionnaire, all participants agreed that the questionnaire was clearly formulated

and well understandable. Additionally, the indicators' validity and reliability were tested by applying the evaluation procedure recommended in Hair et al. [34].

As the pretest showed good results, the measurement items were afterwards included in the questionnaire of the main study (see Table 1).

Table 1. Measurement Items of the Applied Constructs

Perceived Organizational IT-Capability (POITC) *(self-developed)*
from 1='poorer than most' to 7='superior to most'
In relation to other medium-sized companies in this industry, I have the following perception of company A/B/C/D: (POITC1) Overall, the company's IT-capability is… (POITC2) The professional and target-oriented use of IT is… (POITC3) The capability to integrate IT in products and processes is…
Individual IT-Capability (IITC) *(self-developed)*
from 1='poorer than most' to 7='superior to most'
In relation to other students in economic degree programs… (IITC1) I evaluate my overall IT-capability… (IITC2) I evaluate my professional and target-oriented use of IT… (IITC3) I evaluate my capability to integrate IT in products and processes…
Intention to Apply (INTA) [based on 10, 24, 33]
from 1='strongly disagree' to 7='strongly agree'
I evaluate the following statements with regard to company A/B/C/D: (INTA1) If I saw a job opening for this company after graduation, I would apply for it. (INTA2) If I were searching for a job, I would apply to this organization. (INTA3) I would be pleased to work for this organization after my graduation.

4 Data Analysis and Results

The research model was evaluated by applying the PLS-SEM method – implemented in SmartPLS [35] – which is a widely applied method to analyze survey data. It is especially suitable to meet the exploratory nature of this study and works well with small sample sizes [34, 36]. Therefore, it seems a better choice for this study than a covariance-based approach.

4.1 Measurement Validation

The validation of the measurement model followed the established procedure recommended in Hair et al. [34].

To evaluate **internal consistency reliability** the composite reliability value (CR) and Cronbach's α (Cr. α) were computed for each reflectively measured construct. With Cronbach's α >0.7 and CR >0.7 all constructs meet the proposed quality criteria [34, 37] (see Table 2). **Indicator reliability** demands that each indicator's outer loading exceeds 0.7 [34]. With values between 0.816 and 0.970, indicator reliability is also given for each indicator. **Convergent Validity** is ensured with an Average Variance Extracted (AVE) higher than 0.5 [34] for each construct. Therefore, Convergent Validity is also fulfilled (see Table 2). To assess **Discriminant Validity**, the hetereotrait-monotrait ration (HTMT) of the correlations was computed as proposed in Hair et al. [34] and Henseler et al. [38]. As Henseler et al. [38] propose a limit of 0.85 for HTMT values, the HTMT criterion is perfectly fulfilled in this study (see Table 2, bolded values). In summary, the evaluation of the measurement model showed that all constructs and indicators applied in this study meet the established quality criteria for reflective measurement models.

Table 2. Cronbach's Alpha, CR, AVE and HTMT

Construct	*Cr. α*	*CR*	*AVE*	*IITC*	*INTA*
IITC	.863	.915	.782		
INTA	.967	.978	.938	**.323**	
POITC	.939	.961	.891	**.035**	**.390**

4.2 Hypotheses Testing

The evaluation of the structural model including hypotheses testing also followed the established evaluation procedure recommended in Hair et al. [34]: The **structural model** was evaluated for collinearity issues to avoid critical levels of collinearity among the predictor constructs. The VIF values for POITC (1.003) and for IITC (1.001) are both clearly lower than the proposed threshold of 5 [34]. Thus, collinearity is not a critical issue. The amount of **variance explained** in the endogenous variable INTA was 25.7%, which is a high value for behavioral studies [34]. Predictive relevance was assessed using the cross-validated redundancy Q^2 [39, 40]. Q^2 is 0.221 for the sole endogenous variable *Intention to Apply* and therefore indicates predictive relevance ($Q^2>0$ is recommended in Chin [41]).

To evaluate the **significance of the proposed hypotheses** the path coefficients and t-values of the research model were computed applying the PLS Algorithm and PLS SEM Bootstrapping Routine with 5000 subsamples. All hypothesized paths turned out to be significant at a 1% level. The hypothesized direction of the paths was confirmed (see Figure 1). The link between *Perceived Organizational IT-Capability* and *Intention to Apply* (the so-called 'Simple Link') had a path coefficient of 0.379 and a t-value of 6.970 (>2.57) for a two-tailed test. The moderating path had a path coefficient of 0.153 and a t-value of 2.750 (>2.57) for a two-tailed test. The f^2 value, as an indicator of effect size, was 0.193 (>0.15) for the link between POITC and INTA, which is classified as a medium effect [42]. The moderation has an effect size f^2 of 0.038 (>0.025), which indicates a large moderation effect [43].

As the research design may include the risk of not entirely independent observations (see section 3.1), an additional test was conducted to exclude the threat of biased results: The research model was additionally tested with a reduced data set. This reduced set contained only one, randomly selected observation per participant.

The evaluation of the model based on this reduced data set showed very similar results as for the main evaluation. All hypotheses were supported in this analysis as well.

5 Discussion and Implications and Conclusions

This study posed two research questions:

1. Does the perception of organizational IT-capability influence job seekers' intention to apply?
2. Is this influence of particular strength for job seekers with a high level of individual IT-capability?

With regard to the first question, a clear link between the *Perceived Organizational IT-Capability* and job seekers' *Intention to Apply* was found. With regard to the second question, the job seekers' *Individual IT-Capability* turned out to significantly influence the strength of this relationship. Consequently, the perception of a potential employer's IT-capability affects job seekers with a high level of individual IT-capability significantly stronger than job seekers with a low level of individual IT-capability. These findings are consistent with the proposed hypotheses based on implications from prior research.

The study responds to a call for more research with regard to the attraction of proper employees in the initial stage of recruitment [16, 21, 44], especially with a focus on the influence of individual characteristics [12]. It provides initial empirical evidence, how the perception of a specific competition-critical organizational property influences job seekers' application intentions and fosters applicants' self-selection behavior. This knowledge is – in a broader sense – fundamental to better understand the outstanding role of organizational IT-capability in the current 'war for talents'. It provides theoretical and managerial implications as well:

From a theoretical point of view, this research extends the 'traditional' IT-capability perspective, anchored in information systems research, by transferring the common conceptualization of organizational IT-capability (1) to an outsider's perspective and (2) to an individual level. It advances theory and measurement by implementing and validating the concepts of *Perceived Organizational IT-Capability* and *Individual IT-Capability*. These views on IT-capability may provide a proper base for future studies in this context.

As this study bases on 'pure' perceptions of organizational IT-capability without providing any 'hard' data of the organizations' resources and capabilities, it clearly shows that perceptions matter for job seekers' application intentions. In this context, it is in line with (1) prior Marketing-related Research, which showed that the value of a firm's products, resources, and capabilities depends on the perception and appreciation of the customer and with (2) prior Human Resource Research considering the influence of positive perceptions on job seekers' application intentions. It shifts, however, the focus from rather general positive perceptions (e.g., an employer's corporate image) to a specific competition-critical organizational property – an employer's IT-capability. Therefore, this study identified an additional specific driver of job seekers' application intentions. In times of increasing digitization, this 'new' driver will probably continue to grow in importance.

Furthermore, this study helps to better understand the basic mechanisms of applicants' self-selection behavior by combining an applicant's view on an organization's IT-capability and his individual IT-Capability. It clearly shows that individual characteristics affect the strength of the identified relationship between *Perceived Organizational IT-Capability* and *Intention to Apply*. As the question "whether certain individual characteristics or individual differences may have a direct effect on attraction" is poorly investigated so far [12], this study may considerably contribute to this under-investigated field of recruitment research.

As the study provides a deeper understanding of applicants' self-selection behavior especially with regard to the recruitment of IT-specialists, it may help **from a managerial perspective** employers (1) to develop targeted recruitment strategies and (2) to prevent the emergence of vicious circles. Both contributions are of high practical relevance in the current 'War for IT-Specialists':

As the attraction of proper applicants is essential for companies but the goal of attracting the greatest possible number of applicants is questionable [27, 45], employers have to develop an understanding of applicants self-selection mechanisms to particularly address employees with profiles matching the organizations' requirements [44, 45]. In this context, this study clearly showed that information about an organization's IT-capability particularly influences the application intentions of IT-affine job seekers. Keeping all other factors (job description, salary, etc.) constant, they clearly prefer organizations with a (perceived) high level of organizational IT-capability and clearly eschew low-capability organizations, whereas this influence is weaker for less savvy job seekers. Consequently, organizations should clearly promote their IT-capability in recruitment campaigns to attract IT-affine applicants.

This promoting effect of IT-related information in e.g., job ads is of particular relevance for organizations with a rather low level of organizational IT-capability. These firms face the risk to be trapped in a vicious circle, if they do not manage to attract IT-affine employees. They may, however, take advantage from the fact that there is a difference between their objectively existing organizational IT-Capability and applicants' subjective perceptions of this IT-Capability. As these two issues may fundamentally differ, companies can (to some extent) shape job seekers' perceptions of their IT-Capability by providing appropriate information [12]. For example, previous research indicated that the pure amount and specificity of IT-related information may positively influence job seekers' perception development [16].

As the study is an exploratory approach to provide first insights into how *Perceived Organizational IT-Capability* affects job seekers' *Intention to Apply* and IT-affine applicants' self-selection behavior, it has a few **limitations** as well:

First, the study focusses on the perception of organizational IT-Capability. The perception of an organization's IT-capability, however, is only one piece of the puzzle to fully understand job seekers' application intentions. Future research should extend this research and examine – in a comprehensive view – how different drivers of intention to apply interact and if their influence depends on the individual characteristics of a job seeker.

Second, the focus of this study was on advanced students who stated to possibly apply for a job within the next twelve months. As life-circumstances and presumed job opportunities may be different for more experienced job seekers, this study needs to be repeated with a respective sample to investigate parallels and deviations.

Finally, this study bases on hypothetical job ads without giving any 'hard' facts about the respective organizations' IT-Capabilities. It focusses on 'pure' perceptions and questions these perceptions and the resulting behavioral intentions at a single point in time. Ehrhart and Ziegert [12], however, pointed out that the attraction to an organization is a dynamic process. Once attracted by an organization, job seekers are likely to seek additional information to support or adjust their original evaluation [12]. This multi-stage process may gradually diminish the relevance of the job ad information in a real-world scenario.

Appendix

The following job and company descriptions were provided in the study:
The following section provides four extracts of job ads each announcing a job of an internal IT-project manager in a medium-sized manufacturing company with approx. 3000 employees. The standard starting salary is 60,000 Euros. The job profile is characterized as follows for all job ads:

IT-Project Manager As a project manager you are responsible for the successful execution of our internal IT-Projects. You serve as an interface between management and software developer and take responsibility for the further development of our internal IT-Landscape. You are particularly responsible for the steering of the projects with regard to defined objectives, the implementation of development processes, as well as the alignment of IT-Projects and the company's business strategy.

The job is announced by four different companies which differ with regard to their company description. The four companies provide the following information in their job ads:

Company A: We value Information Technology as a foundation of an efficient management and the continuous evolution of our company. We pay special attention to benefit from the opportunities of Information Technology to continuously develop our business value. We pursue a clear IT-strategy to reach our goals. Our IT team keeps its finger on the pulse by applying agile methods and the latest technologies. We see our strength particularly in the area of data management and mobile applications. We continuously seek new ways to develop and deploy our IT-landscape even more effective.

Company B:
As a medium-sized manufacturing company, we traditionally focus on the sector of classical engineering. As Information Technology is continuously gaining influence on the efficient organization of our internal processes as well as on our supply chain, we aim at continuously developing our rather traditional IT-landscape. We are looking for a creative employee to develop and implement proper IT-solutions and encourage innovative suggestions with regard to flexible IT-architectures and data structures.

Company C:
We are a regionally rooted company with a long tradition in the manufacturing industry. We combine tradition with the innovative opportunities of Information Technology. We particularly focus on the storage and usage of data to support our supply chain (e.g., Shared Data Base) and as a basis of interconnected ERP-Systems. Our special concern is to view IT as an integrative part of our business strategy and to identify and efficiently deploy economically viable IT projects. As we are continuously looking for new ways, we provide an open and supportive environment for new ideas.

Company D:
Our company bundles the know-how from purchasing, logistics, and supply-chain management in interdisciplinary teams aiming at an efficient manufacturing of our products. In this context, Information Technology is an important component to develop this efficiency. We focus on the cross-linking of business processes, data bases and information systems and are already applying agile methods of process and project management. We are open-minded towards innovative, data-driven extensions of our existing business models and strive for further optimizations of our supply chain processes.

References

1. Chen, Y., Wang, Y., Nevo, S., Jin, J., Wang, L., Chow, W.S.: IT capability and organizational performance: the roles of business process agility and environmental factors. European Journal of Information Systems. 23, 326–342 (2014).
2. Lu, Y., Ramamurthy, K.: Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination. MIS Quarterly. 35, 931–954 (2011).
3. Wade, M., Hulland, J.: The resource-based view and information systems research: Review, extension, and suggestions for future research. MIS quarterly. 28, 107–142 (2004).
4. Bharadwaj, A.S.: A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. MIS Quarterly. 24, 169–196 (2000).
5. Stoel, M.D., Muhanna, W.A.: IT capabilities and firm performance: A contingency analysis of the role of industry and IT capability type. Information and Management. 46, 181–189 (2009).

6. Spitzer, B., Morel, V., Buvat, J., Kvj, S.: The digital talent gap: Developing skills for today's digital organizations. Capgemini Consulting. (2013).
7. Cable, D.M., Judge, T.A.: Pay preferences and job search decisions: A person-organization fit perspective. Personnel psychology. 47, 317–348 (1994).
8. Cable, D.M., Judge, T.A.: Person–organization fit, job choice decisions, and organizational entry. Organizational behavior and human decision processes. 67, 294–311 (1996).
9. Kristof, A.L.: Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. Personnel psychology. 49, 1–49 (1996).
10. Wei, Y.-C., Chang, C.-C., Lin, L.-Y., Liang, S.-C.: A fit perspective approach in linking corporate image and intention-to-apply. Journal of Business Research. 69, 2220–2225 (2016).
11. Wolgast, S., Björklund, F., Bäckström, M.: Applicant Ethnicity Affects Which Questions Are Asked in a Job Interview. Journal of Personnel Psychology. 17, 66–74 (2018).
12. Ehrhart, K.H., Ziegert, J.C.: Why Are Individuals Attracted to Organizations? Journal of Management. 31, 901–919 (2005).
13. Sánchez-Fernández, R., Iniesta-Bonillo, M.Á.: The concept of perceived value: a systematic review of the research. Marketing Theory. 7, 427–451 (2007).
14. Grönroos, C.: Value-driven relational marketing: From products to resources and competencies. Journal of Marketing Management. 13, 407–419 (1997).
15. Nwankpa, J.K., Roumani, Y.: IT Capability and Digital Transformation: A Firm Performance Perspective. In: Proceedings of the 37th International Conference on Information Systems. pp. 1–16. , Dublin (2016).
16. Roberson, Q.M., Collins, C.J., Oreg, S.: The Effects of Recruitment Message Specificity on Applicant Attraction to Organizations. J Bus Psychol. 19, 319–339 (2005).
17. Turban, D.B.: Organizational attractiveness as an employer on college campuses: An examination of the applicant population. Journal of Vocational Behavior. 58, 293–312 (2001).
18. Hogg, M.A., Terry, D.I.: Social identity and self-categorization processes in organizational contexts. Academy of management review. 25, 121–140 (2000).
19. Beach, L.R.: Image Theory: Decision Making in Personal and Organizational Contexts. John Wiley & Sons Ltd, Chichester, West Sussex, England ; New York (1990).
20. Stevens, C.K.: Image Theory and Career-Related Decisions: Finding and Selecting Occupations and Jobs. Image theory: Theoretical and empirical foundations. 227–239 (1998).
21. Barber, A.E.: Recruiting Employees: Individual and Organizational Perspectives. SAGE Publications, Thousand Oaks, London, New Dehli (1998).
22. Cable, D.M., Turban, D.B.: Establishing the dimensions, sources and value of job seekers' employer knowledge during recruitment. Research in personnel and human resources management. 20, 115–163 (2001).
23. Cable, D.M., Turban, D.B.: The value of organizational reputation in the recruitment context: A brand-equity perspective. Journal of Applied Social Psychology. 33, 2244–2266 (2003).
24. Collins, C.J.: The interactive effects of recruitment practices and product awareness on job seekers' employer knowledge and application behaviors. Journal of applied psychology. 92, 180 (2007).
25. Turban, D.B., Cable, D.M.: Firm reputation and applicant pool characteristics. Journal of Organizational Behavior. 24, 733–751 (2003).
26. Gatewood, R.D., Gowan, M.A., Lautenschlager, G.J.: Corporate image, recruitment image and initial job choice decisions. Academy of Management journal. 36, 414–427 (1993).

27. Rynes, S.L., Bretz, R.D., Gerhart, B.: The importance of recruitment in job choice: A different way of looking. Personnel psychology. 44, 487–521 (1991).
28. Steenkamp, J.-B.E.: Conceptual model of the quality perception process. Journal of Business research. 21, 309–333 (1990).
29. Bretz, R.D., Judge, T.A.: The role of human resource systems in job applicant decision processes. Journal of Management. 20, 531–551 (1994).
30. Judge, T.A., Bretz, R.D.: Effects of work values on job choice decisions. Journal of applied psychology. 77, 261–271 (1992).
31. Judge, T.A., Cable, D.M.: Applicant personality, organizational culture, and organization attraction. Personnel psychology. 50, 359–394 (1997).
32. Brislin, R.W.: Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology. 1, 185–216 (1970).
33. Taylor, M.S., Bergmann, T.J.: Organizational recruitment activities and applicants' reactions at different stages of the recruitment process. Personnel Psychology. 40, 261–285 (1987).
34. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modeling. Sage Publications Ltd., Los Angeles (2016).
35. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. Boenningstedt: SmartPLS GmbH (2015).
36. Fornell, C., Bookstein, F.L.: Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing research. 440–452 (1982).
37. Bagozzi, R.P., Yi, Y.: Specification, evaluation, and interpretation of structural equation models. Journal of the Academy of Marketing Science. 40, 8–34 (2012).
38. Henseler, J., Ringle, C., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science. 43, 115–135 (2015).
39. Geisser, S.: The predictive sample reuse method with applications. Journal of the American statistical Association. 70, 320–328 (1975).
40. Stone, M.: Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological). 111–147 (1974).
41. Chin, W.W.: How to write up and report PLS analyses. In: Esposito Vinzi, V., Chin, W.W., Henseler, J., Wang H. (eds.) Handbook of partial least squares. pp. 655–690. Springer, Berlin, Heidelberg (2010).
42. Cohen, J.: Statistical power analysis for the behavioral sciences . Hilsdale. NJ: Lawrence Earlbaum Associates. 2, (1988).
43. Kenny, D.A.: Moderation, http://davidakenny.net/cm/moderation.htm. (Accessed: 17.07.2018)
44. Gomes, D., Neves, J.: Organizational attractiveness and prospective applicants' intentions to apply. Personnel Review. 40, 684–699 (2011).
45. Breaugh, J.A., Starke, M.: Research on employee recruitment: So many studies, so many remaining questions. Journal of management. 26, 405–434 (2000).

Good, Bad, or Both? Conceptualization and Measurement of Ambivalent User Attitudes Towards AI

Sophia Bettina Maier[1], Ekaterina Jussupow[1], and Armin Heinzl[1]

[1] University of Mannheim, General Management and Information Systems, Mannheim, Germany
sophia.b.maier@googlemail.com
Corresponding author: jussupow@uni-mannheim.de
heinzl@uni-mannheim.de

Abstract. Artificial intelligence is currently one of the most controversial discussed technologies across various work domains. In healthcare, AI as decision aid fosters physicians' fears of being replaced and strong positive beliefs of increasing quality of care. This causes ambivalence as individuals hold both strong positive and negative attitudes simultaneously. Yet, most studies on technology adoption conceptualize attitudes as unidimensional and do not capture ambivalent attitudes. Furthermore, there is a lack of quantitative studies on ambivalence and thereby a lack of insights on characteristics and impact of ambivalent attitudes. Our study measures different states of ambivalence in the context of healthcare and links it to resistance to change. We are able to show that ambivalent attitudes do cause resistance to change when individuals who hold both strong positive and negative evaluations also experience an inner conflict (felt ambivalence). Furthermore, we are able to identify distinct context-specific attitudes which foster physicians' ambivalence.

Keywords: Ambivalence, User Attitudes, Resistance, Artificial Intelligence, Healthcare

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

A Case Study on Cross-Hierarchical Communication in Digital Work Environments

Christian Meske[1], Tobias Kissmer[2], Stefan Stieglitz[2]

[1] Freie Universität Berlin and Einstein Center Digital Future, Berlin, Germany
{christian.meske}@fu-berlin.de
[2] University of Duisburg-Essen, Department of Computer Science and Applied Cognitive Science, Duisburg, Germany
{tobias.kissmer,stefan.stieglitz}@uni-due.de

Abstract. Hierarchies are vital for organizations. However, in the workplace they can also act as invisible barriers for the communication between individuals from different hierarchical levels, thus impeding the exchange of information and knowledge. This effect is particularly assumed to be true for country cultures with high power distance characteristics. Based on a unique Skype for Business dataset, we test the potential impact of country culture for cross-hierarchical communication. Finally, we aggregate the data of all countries and provide insights into specific communication patterns for each hierarchical level across the organization. We conclude our case study by providing distinct lessons learned, such as the irrelevance of cultural predictions in this context and the essential focus on the middle levels of hierarchy.

Keywords: Digital Workplace, Hierarchy, Power Distance, Culture

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

‘Show Me Your People Skills’ - Employing CEO Branding for Corporate Reputation Management in Social Media

Milad Mirbabaie[1], Julian Marx[1], and Stefan Stieglitz[1]

[1] University of Duisburg-Essen, Department of Computer Science and Applied Cognitive Science, Duisburg, Germany
{milad.mirbabaie, julian.marx, stefan.stieglitz} @uni-due.de

Abstract. Digitization has led to increased social media utilization among companies to connect with their customers. We know that particularly CEOs, as the representing face of a company, can exert great influence to build corporate reputation. While reputation management in general has been researched extensively, we know little about the dimensions of CEO reputation management in social media. This paper deals with the distinction of organizational and personal branding in Twitter, and moreover, aims to determine eligible dimensions for CEO reputation management, based on the widely accepted Reputation Quotient. Therefore, we collected 3,604 social media postings of companies and their respective CEOs from Twitter. Through statistical and content analyses, we determined two supplementary dimensions for CEO reputation. *Shared Interests* and *Personal Logging* add private aspects to the spectrum of CEO reputation management, which have the capacity to foster consumer engagement.

Keywords: Social Media, Communication Strategy, Reputation Management, Organizational Branding, Personal Branding

1 Introduction

In recent years, social media has become an indispensable tool for brand communication. Compared to traditional marketing channels, social media provides a cost-effective platform for companies to promote products and interact with consumers [1], [2]. Since the digitization has progressed, a company is no longer judged solely on the basis of its products and quality of services, but also in terms of its communication and interaction with consumers. Customers want to speak up, provide feedback, ask questions and receive supplementary product information [3]. In addition to organizational brand presentation, top managers such as chief executive officers (CEOs) use social media pages to provide a platform for interaction and exchange [4]. CEOs, as the public face of a company, may involve private aspects in their posting

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

routine, and thus be perceived as more accessible and authentic [5]. As a result, not only boundaries between personal and professional social media use blur, but companies may also capitalize on the influence that personal brands exert on consumers [6]. While research acknowledges the impact of CEO personal branding on the reputation of the organization [7–9], little is known about how to conceptualize CEO reputation management in social media. Despite the consensus on the potential impact of a strong CEO brand page, few companies pursue it as part of their communication strategy.

Our study aims to address this issue by investigating the role of a CEOs Twitter presence alongside the corporate branding and motivate practitioners to intensify their efforts to employ a strong CEO branding in order to complement their reputation management in social media. The dimensions that matter, on this account, are the ones that occur when organizational communication is carried out through the individual channel of a CEO. On this basis, our paper centers around following research questions:

RQ1: *How do communication strategies of CEO and corporate branding differ in Twitter communication?*
RQ2: *What dimensions does CEO branding in social media add to corporate reputation management?*

Findings based on the above research questions would significantly enhance our understanding of CEO reputation management in social media. The contribution of this paper is primarily of practical nature, i.e. an improved utilization of CEO personal branding to build up a reputation that is detached from the organizational brand and is, for exactly this reason, valuable to the organization. At the same time, however, our study attempts to conceptualize CEO reputation management with consideration of Twitter-specific capacities such as retweet and follower mechanics. To address both research questions, we tracked social media data from Twitter over a timespan of eight weeks, resulting in a collection of 3,604 tweets that were expressed by leading firms within the IT industry and their respective CEOs. Our research methodology encompasses statistical analyses to determine how well tweets were perceived and content analyses to advance toward a better understanding of CEO reputation management on Twitter from a qualitative perspective. In order develop a predictive model for CEO reputation, which can be tested, this study first targets to establish a conceptual basis by means of an explorative approach. This offers promising opportunities for supplementary quantitative research on this matter. From a practical point of view, communication and IT strategists, and CEOs of large and medium-sized enterprises benefit from this study. Aligning the textual dimensions of CEO communication with capacities of social media is crucial for effective reputation management.

The paper is structured as follows. In a primary step, we review previous literature and form a theoretical background to this study. Subsequently, we outline our research design including methods and data analysis measure. We then present our findings and theorize relevant outcomes toward a concept of CEO reputation management in social media. Finally, we conclude our study with a summary, implications, limitations, and suggest further research prospects.

2 Related Work

With the prevalence of social media for brand communication, online reputation management as a relevant factor of influence for firm value and performance came into scholarly focus [10–12]. In comparison to traditional computer-mediated communication, social media use impacts organizational communication processes as they “afford behaviors that were difficult or impossible to achieve in combination before these new technologies entered the workplace” [13, p. 143].

Since corporate reputation is based on various pillars regarding the perceptions of stakeholders, companies unanimously try to govern associations with their organization toward establishing a favorable image [12], [14–17].

One of the main drivers for organizations to invest in online activities and engaging with consumers is the positive impact on corporate reputation [10]. Since consumers prefer companies with a positive reputation, the efficient use of social media platforms can promote customer loyalty and retention [12]. Furthermore, it may sustain a company's performance to consequently translate into higher market value [17–19]. Even in times of a crisis, a beneficial reputation can protect a company from negative consequences [21].

Through social media, however, corporate reputation has become increasingly determined by outside forces. Social media coverage is less manageable through public relations due to the empowerment of the consumer to publicly evaluate a company’s reputation and influence others [22]. Some scholars have observed the challenge for companies to appear as trusted actors on social media that have the capability to actively manage their reputation [10]. One way to tackle this issue is to split the inflows of corporate reputation on separate shoulders and complement a firm’s communication strategy with individual branding of high-ranking executives. CEOs, in particular, have the ability to personalize the organization, and at best, be a trusted actor. Not to mention that a strong CEO brand helps to establish positive relations with a variety of stakeholders [23], [24].

The association with a positively perceived public face may reflect upon the organization as a whole [25]. Literature, in this context, picked up the notion of ‘social CEOs’, which characterizes executives as socially acclaimed mediators who bring their organization closer to their customers [26]. CEOs, too, face affordances of social media platforms and applications that respond to their psychological needs [27]. As a matter of fact, they are perceived as the incarnation of corporate identity by an different groups of stakeholders [25]. Therefore, it is imperative for organizations to systematically incorporate CEO reputation in communication processes.

Existing literature provides evidence for examining reputation management on the basis of Twitter data. For instance, a study by Capriotti and Ruesja [4] analyzed the presence, activity, and interaction of CEOs on Twitter and found that the reputation of a company’s CEO affects and may even define the image of a company. In addition, CEOs may also positively affect their own career [7]. However, common practices of CEOs on Twitter oftentimes do not align with the character of the medium, e.g. engaging in one-sided conversations on a two-way platform [28]. Thus, considerable potential remains untapped when it comes to develop effective social media strategies for a CEOs reputation management. Moreover, Weng and Chen [29] conducted a study on the impact of CEO and corporate reputation on financial performance to compare the two. Both corporate and CEO reputation entail separate effects on the financial performance, but CEO reputation is considered as more relevant. Even if corporate reputation is observed as poor, the reputation of the respective CEO may still exert a positive impact on the financial performance.

Literature in this context does not comprise a sufficient framework that is applicable to CEO reputation management in social media. Therefore, in this study, we turn to fundamental work on corporate reputation management to be able to conceptualize our findings on CEO reputation management.

3 The Reputation Quotient

As a measurement concept to determine reputation and to differentiate between single dimensions of it, Fombrun et al. [30] introduced the Reputation Quotient (RQ). It has been acknowledged as a well-accepted basis for quantitative measurements regarding organizational reputation [31], [32]. Numerous studies have been conducted to apply and improve the RQ in practice, even in cross-cultural settings [31], [33]. Thus, we endorse this stream of research and employ the RQ as an eligible classification scheme to identify social media users' perceptions about the reputation of a company. To support this study, the suggested dimensions of corporate reputation serve as a template to be synchronized with CEO reputation. This allows to determine whether CEO reputation management can be assessed through existing reputational dimensions, or if additional dimensions are required for a conceptualization of CEO reputation management. As shown in figure 1, the RQ subdivides corporate reputation into six dimensions: *Emotional Appeal*, *Products and Services*, *Vision and Leadership*, *Workplace Environment*, *Social and Environmental Responsibility,* and *Financial Performance*.

The dimensions were established as a result of interviewing focus groups by means of a reputation survey. Hence, the dimensions reflect how a company is perceived from various angles. Benchmarking studies can provide important contributions here by analyzing which companies were able to achieve which reputation values with the help of which measures. This may include extended variants of the RQ detached from the measurement procedure of the initial RQ study [34]. To serve as a benchmark for our study, we employ the existing RQ dimensions to determine how corporate reputation management can proactively be addressed, i.e. tailoring social media content to match relevant dimensions.

Including *Emotional Appeal* in a communication strategy may result in positive feelings and respect for the company and eventually increase trust. Covering *Products and Services* is rather oriented toward marketing and holds the capacity to lead consumers to perceiving an organization as innovative, expecting high product quality, or to express identification with those products and services. *Vision and Leadership* conveys the corporate mission as well as a goal-oriented execution of a company's activities. Moreover, proficient management and the impression of an organization to be a popular workplace is part of the *Workplace Environment* dimension. *Social/Environmental Responsibility* aims to reflect a firm's commitment to good causes and responsibility towards environment and society. Finally, *Financial Performance* indicates an organization to be profitable, and to be capable to outperform competitors. Therefore, it is expected to continue to grow in the future and please shareholders [32].

The dimensions suggested by the RQ have been applied successfully for research in social media settings based on messages regarding corporate reputation [12], [35]. The RQ might serve as a useful baseline as CEO reputation, to a large extent, aligns with corporate reputation [24], [36]. However, as the taxonomy of the RQ was developed in a distinct context, it remains unclear if it seamlessly applies to CEO reputation management. We evaluate the RQ's usefulness in this matter by assessing its efficacy in classifying a random sample of CEO communication. According to the guidelines of [37] for taxonomy development, we define the professional leitmotif to be the meta-characteristic of each tweet. To evaluate the usefulness of the RQ dimensions, we square the data provided by our content analysis with the characteristics of each RQ dimension, and if necessary, expand the RQ typology to get closer to a general taxonomy for CEO reputation.

4 Research Design

4.1 Data Collection and Preprocessing

To obtain relevant data, we collected publicly accessible social media postings from Twitter. Due to its velocity and publicness [38], [39], Twitter serves as an eligible data source for this research endeavor. We arrayed a preliminary sample of firms from a pool of the *"Top 50 Global Technology Companies"* published by Fortune Magazine [40], which is based on overall market share. Moreover, we limited our selection of firms to the US market due to a high affinity for professional Twitter usage in the United States. We further established the criteria that each company and its respective CEO had to be active (≥5 original tweets per month) on Twitter prior to the tracking. Gathering data of both the company and the CEO allows a direct comparison of prevailing dimensions of reputation management. In order to evaluate CEO reputation management, we need the social media presence of the corporate brand to serve as a benchmark. Our tracking encompassed the activity of following accounts: @twitter, @amazon, @google, @apple, @microsoft, @intel, @tesla, @xbox, @jack, @JeffBezos, @satyanadella, @sundarpichai, @tim_cook, @bkrunner, @elonmusk, @XboxP3.

By means of a self-developed Java crawler and the open source library Twitter4J, we captured a total of sixty-one days of Twitter communication from April 25th, 2018 (0:00 UTC) to June 24th, 2018 (23:59 UTC). The crawler was set to only gather data provided with English language settings. We identified the official Twitter accounts of our company and CEO pairs and tracked all their Twitter activities. Extracted data was stored in a MySQL database for further preprocessing. We finally exported tables of the complete sample and used Tableau and Microsoft Excel to perform our analyses.

4.2 Statistical Analysis

In order to get a deeper sense of our dataset, we provide basic statistical metrics, which are primarily Twitter-specific. This includes the retrieval of metrics such as the follower count of the authors, but also key figures of single postings such as the retweet and favorite count of each posting [41–43]. Moreover, we calculate different ratios to ensure the comparability of the research subjects. To evaluate each dimension with regard to its impact on reputation management, we need to consider the preconditions of each account. Literature suggests that the follower count of Twitter users indicates its popularity, and therefore its potential reach [44]. Furthermore, we know that the number of retweets are suggestive of a messages' influence [43]. Hence, we consider a relational index of retweets per thousand followers ($R = \frac{r}{fw \times 0{,}001}$) or favorites per follower ($F = \frac{f}{fw \times 0{,}001}$). These values point to an assertion how tweets addressing certain RQ dimensions are perceived within the Twitter community.

4.3 Content Analysis

To develop a typology of CEO communication in social media we utilize inductive category formation based on the method of qualitative content analysis by Mayring [45]. Here, we reduce Twitter messages to their main point and sort them into categories. The coding process is illustrated in figure 1.

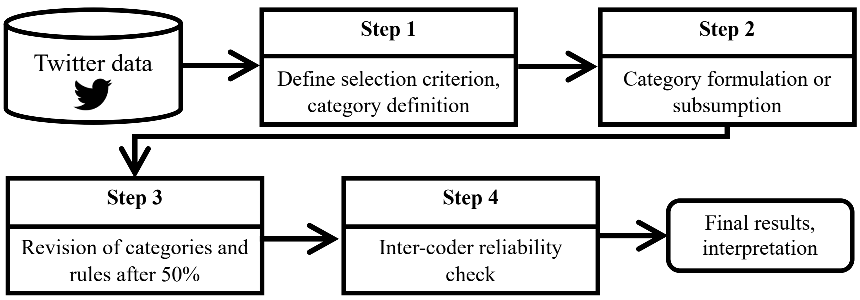


Figure 1 Category development and sample coding procedure, based on [38].

We define both text-based content as well as supplementary audiovisual content of tweets and retweets as the selection criteria. Our initial set of categories consists of the reputational dimensions provided by the RQ (step 1). In case conveyed messages are not covered by the RQ dimensions, e.g. "Wow the sky in Boston right now is crazy. Orange!", we create an additional category. In this case, the message and added photograph qualify the message to be allocated to the new category 'personal logging' (step 2). After coding 50% of the data line by line, we revise the set of categories to make sure it covers the whole spectrum of message types (step 3). The entire sample is set to be manually coded by each author to ensure inter-coder reliability (step 4). Using Krippedorff's alpha, a score of .827 was calculated. Our coding can be rated as reliable as $\alpha \geq .800$ [46]. We examine the sample in its entirety to discover commonalities and topical patterns. This will serve as a basis to answer our second research question and to theorize our findings with regard to CEO reputation management.

5 Results

Our final sample consists of eight companies and their corresponding CEO: *Twitter/Jack Dorsey, Amazon.com/Jeff Bezos, Intel/Brian Krzanich, Google/Sundar Pichai, Microsoft/Satya Nadella, Tesla/Elon Musk, Apple/Tim Cook, and Xbox/Phil Spencer.* Within the 8-week period of our data tracking, those accounts authored a total of 8,628 tweets. Figure 2 illustrates how those tweets spread among tracked accounts and how often they retweeted others.

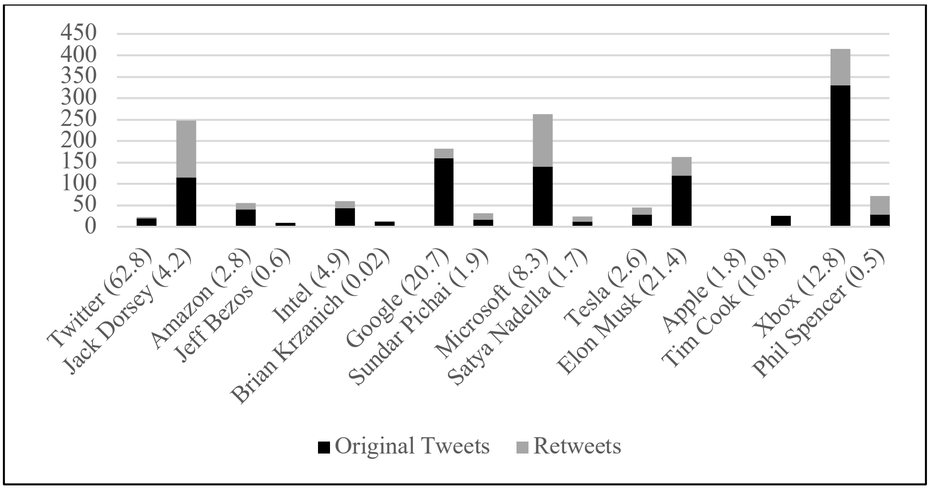


Figure 2 Tweet activity and follower count (in parentheses), April 25th 2018 - June 24th 2018

The data includes 1,101 original postings, 526 retweets, and 7,001 @mentions. The latter directly address a specific user but do not show up in the timeline of both communicator's followers. The large number of mentions primarily stem from @Xbox (4,137) and @Google (1,981). We were not able to capture tweets from the @apple account as it is solely use for nonrecurring advertising. Except for Twitter, CEOs author significantly less tweets compared to their corporate brand account. Those relations align with follower counts, as all investigated corporate brands have larger fan bases than their respective CEO.

Coding all original tweets and retweets resulted in a diverse distribution of addressed RQ dimensions. Figure 3 shows the spectrum of each account in relative proportions.

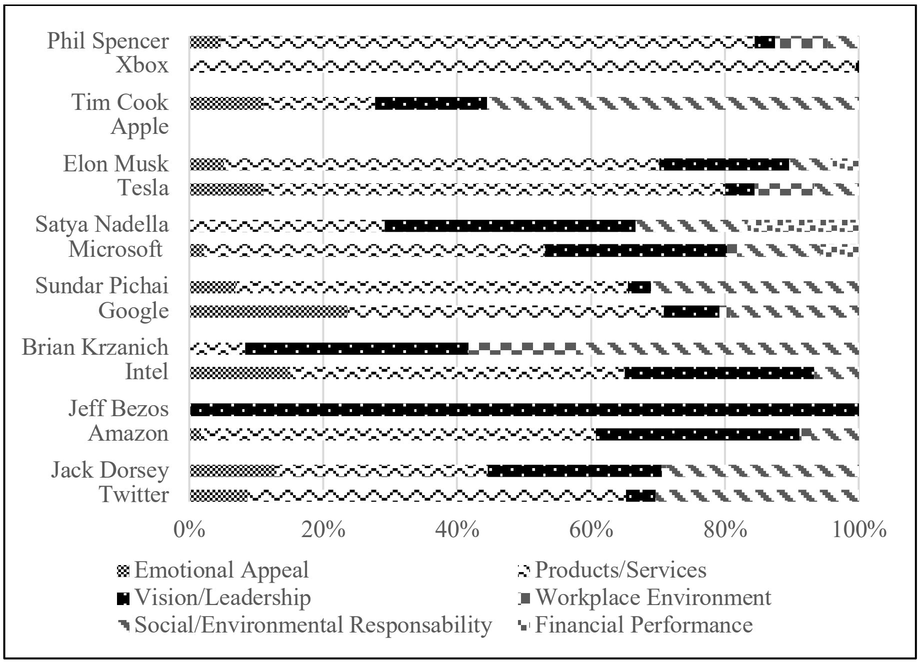


Figure 3 Spectrum of addressed RQ dimensions per company/CEO pair

Among the corporate brands, advertising products and services tends to be the predominant dimension. For CEOs, however, this dimension seems less relevant. While CEOs, too, advertise products and services to some extent, they rather focus on authoring tweets addressing vision and leadership and Social/Environmental Responsibility. Each tweet was exclusively allocated to one category. We ended up with 208 CEO authored tweets that did not fit the description of the initial six categories provided by the RQ dimensions. Consequently, we formulated additional categories to subsume those messages. Table 1 lists the total figures of tweets and retweets being categorized.

Table 1 Coding results (absolute figures) per dimension

	EA	*PS*	*VL*	*WE*	*SE*	*FP*	*SI*	*PL*
Companies	66	748	130	13	87	0		
CEOs	30	175	77	8	78	7	107	101

In a second pass of coding this subsample, we observed that those tweets were exclusively linked to the personality of the CEO. The first subgroup of messages (107) contained *Shared Interests* of the CEO that were of private nature. For example, Jeff Bezos authored a tweet containing a pop culture reference to remember an author who recently passed away. In a second example, Sundar Pichai expressed his expectations for the upcoming FIFA World Cup, naming a few teams he favored. The according tweets are represented in figure 4.

Figure 4 Tweet examples of the Shared Interest dimension

As a second group of messages (107), we identified postings that contained information about personal experiences and/or statements of the CEO. Following the popular term of blogging (and vlogging), we grouped those tweets under the notion of *Personal Logging*. It is important to notice that those tweets are authored from an ego-perspective. This excludes media or PR reports about the CEO. Jack Dorsey, for instance, shared his view from a rooftop in Boston. On Mother's Day, Jeff Bezos shared rather intimate thoughts about his mother, accompanied with a picture of her. The respective tweets are represented in figure 5.

Figure 5 Tweet examples of the Personal Logging dimension

In order to assess above dimensions with regard to their impact on CEO reputation management, we calculated relational indices. Follower counts were retrieved at the beginning of the tracking, and therefore, treated as an invariable value. Figure 6 depicts an overview of such measurements for better clarity.

For the R index (retweets per thousand followers), we calculated the mean value of retweets received for the sum of all original tweets and retweets posted or forwarded by each user. The F index (favorites per thousand followers), the retweets were excluded, as only original tweets collect favorites. We used abbreviations for better readability *(EA=Emotional Appeal, PS=Products/Services, VL=Vision/Leadership, WE=Workplace Environment, SE=Social/Environmental Responsibility, FP=Financial Performance, SI=Shared Interest, PL=Personal Logging)*, and listed our results in table 2.

Table 2 RQ dimension engagement (R=retweets per thousand followers; F=favorites per thousand followers)

CEO	*Follower Count*		*EA*	*PS*	*VL*	*WE*	*SE*	*FP*	*SI*	*PL*
Jack	4,212,972	R	1.95	0.03	0.38	0.01	1.2	0	1.6	0.02
Dorsey		F	0.07	0.13	0.2	0	0.26	0	0.1	0.23
Jeff	590,217	R	0	0	5.24	0	0	0	0.59	2.25
Bezos		F	0	0	18.11	0	0	0	4.39	20.92
Brian	20,778	R	0	1.88	2.11	1.17	0.92	0	0	0
Krzanich		F	0	4.14	5.04	4.93	4	0	0	0
Sundar	1,899,332	R	0.1	0.16	1.86	0	0.51	0	0.1	0
Pichai		F	0.74	1.39	5.41	0	3.64	0	2.39	0
Satya	1,717,000	R	0	0.18	0.13	0	0.34	0.8	0	0
Nadella		F	0	1.2	0.67	0	1.77	3.63	0	0
Elon	21,422,347	R	0.15	0.16	0.47	0	0.15	0.49	0.05	0.35
Musk		F	3.37	2.60	3.10	0	1.49	2.42	0.64	2,74
Tim	10,821,989	R	0.05	0.20	0.74	0	0.14	0	0.03	0.06
Cook		F	0.45	0.89	2.33	0	0.75	0	0.36	0.45
Phil	588,243	R	0.98	1.08	4.29	0.10	0	0.19	0.40	1.26
Spencer		F	0	2.27	2.18	1.04	0.84	0	3.23	8.64

The above findings suggest that the *Vision/Leadership* is prominently being addressed by the CEOs considered in our sample. At the same time, those postings provoked more engagement compared to the other dimensions. *Products/Services* and *Social/Environmental Responsibility* are part of the CEOs communication strategy in 4 out of 5 cases. We can further derive from this examination that Twitter users are more likely to add a tweet to their favorites rather than retweeting it. It further varies how much engagement CEOs receive from their following. For instance, Jeff Bezos received 5.24 retweets per thousand followers when addressing VL, whereas Jack Dorsey received only 0.38 retweets from his visionary statements.

We further observe that the supplementary dimensions of CEO reputation management are more often addressed than primary dimensions, e.g. the financial performance of their company. Considering SI and PL, Jack Dorsey received mediocre to little feedback when addressing personal issues. Jeff Bezos, however, received 20.92 favorites per thousand followers when taking his followers along his personal life. This was the highest score among all values.

6 Discussion

6.1 Distinguishing CEO and Corporate Branding (Research Question 1)

Our initial approach was to find similarities and contrasts of a corporation's communication strategy and the social media presence of its highest representative. In our sample, we found corporate brand pages to fall back on larger followings than their CEO-pendants. Moreover, the publishing activity of corporate brand pages is significantly higher than the one of CEO brand pages. Corporate brand pages may be used to supplement a firm's customer-service (Google, Amazon), which is apparent through intensive use of @mentions. This adds a much more conversational tone to the customer-brand relationship and, in this case, adapts to the character of Twitter.

Nevertheless, the corporate brandings are furnished with much more diligence in terms of content creation and publishing schedule. CEOs, in contrast, publish in a more sporadic and ad-hoc manner. In our study, we observed different strategies how CEOs may complement their corporate brand pages. In most cases, the CEOs followed similar agendas, e.g. tweeting and retweeting about *Products/Services* or the *Social/Environmental Responsibility* of their firm (e.g. Sundar Pichai, Satya Nadella). In other cases, the CEOs communication strategy supplemented their firm's efforts by focusing on different dimensions such as Vision/Leadership (Jeff Bezos, Brian Krzanich) or the Workplace Environment (Brian Kraznich). Measured against their follower base, the latter strategy provoked comparatively high engagement.

Our study displays that reputation management in social media is primarily employed to reach the customer base [12], [14]. Other stakeholder groups are rather neglected, as only Microsoft published tweets concerning the financial performance of their corporation. This situates social media in a unique position as they have the tendency to be channels for immediate action upon the customer's perception of the firm. This means increased empowerment to affect corporate reputation, but also higher volatility in how a company's reputation can change over time [14]. Hence, studies on reputation management in social media may find different dimensions of reputation management to be emphasized than, for instance, the financial performance of a company [32].

From a theoretical perspective, found the RQ dimensions to be eligible tools of classification when it comes to corporate reputation management [30]. However, in order to cover the entire spectrum of a CEOs reputation management, additional dimensions were required to mirror the capacities social media has to offer for CEO branding.

6.2 Supplementary Dimensions for CEO Reputation Management (Research Question 2)

Content analysis measures of social media postings attached to a corporate brand revealed that Fombrun's work including the RQ dimensions are doubtlessly applicable to the shape and form of contemporary online reputation management [16], [17], [30]. At the same time, this predicates that companies tend to not break new ground on a regular basis, but rather transfer established strategies onto new channels. CEOs, however, face the opportunity to meet with customers on a fundamentally different level of communication. five out of eight sampled CEOs portrayed themselves as fathers, husbands, and sports fanatics in front of their audience [5], [26]. Taking personal branding to such levels is a controversial choice to make. It blurs boundaries between private and professional life and leaves a person no choice but to be unhidden on a constant basis.

Despite these concerns, personal branding has become a commercial model of significant magnitude. Our study reveals that CEOs with exemplary roles in the IT industry cautiously discover personal branding elements as part of their image building in social media. Due to the inseparability of CEO and corporate reputation, theories dealing with reputation management require to widen their scope to incorporate the social mutation of CEOs. We argue that CEOs better meet the capacities of social media, as well as psychological needs of customers and themselves when communicating individually on social media [27]. This particularly holds for communication that addresses personal dimensions. Parasocial relationships between customers and CEOs may strengthen not only the CEO's reputation [25],[26], but at the same time, be carried over to the organizational brand.

Our results emphasize blurring boundaries between private and professional competences for CEOs, evoked through the ubiquity of social media and an amended brand-customer relationship. Nevertheless, personal branding may add significant value to the communication strategy of a CEO and result in a competitive advantage due to enhanced reputation. A supplementary individual communication strategy through the CEO brand holds the potential to increase overall engagement in Twitter. However, incorporating personal dimensions in professional brand communication should meet psychological needs of a CEO, who might set individual boundaries of self-disclosure.

Accordingly, we propose to append personal dimensions to our understanding of CEO reputation management. This first explorative investigation spawned *Shared Interests* and *Personal Logging* as relevant dimensions of CEO reputation management in social media, as shown in figure 6.

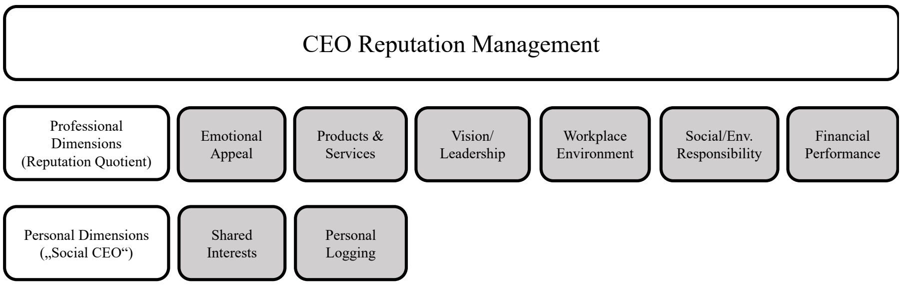


Figure 6 Dimensions of CEO Reputation in social media

The humanization of a corporate brand only works to a certain extend. Even though IT-firms push those boundaries through recent technological advancements such as social bots or conversational agents, an ever-present CEO has significant impact on the public image of a company. While fundamental work of [30] decently covers what we know as corporate reputation, extending the RQ dimensions to consider contemporary challenges and opportunities of CEO reputation management seems to be the next logical step. Hence, we need to allow for the individual characters of high executives to systemize their efforts in order to contribute to their personal and corporate brand, and consequently, to better serve their customer base.

7 Conclusion and Further Research

The study aims to impart new impetus to the landscape of research on reputation management. With regards to the ubiquitous role of social media, we award a new role to CEOs in terms of representing the fortunes of a company in a more accessible manner. By turning to [30] and their Reputation Quotient, we undertook efforts to adopt this set of dimensions to the specifications of CEO reputation management in social media. Through data collection by means of a self-developed java crawler and the official Twitter API, we were able to perform statistical and content analysis measures on a dataset of 8,628 tweets. The analysis of our sample, containing of corporate brands and their corresponding CEOs from the IT sector, revealed different RQ dimensions to be part of each communication strategy. Whereas corporate brands focus on promoting *Products/Services* and expressing their *Social/Environmental Responsibility*, CEOs authored *Vision/Leadership* statements more frequently. Moreover, we identified two additional dimensions, *Shared Interests* and *Personal Logging*, to complement the RQ and make it applicable to CEO reputation management.

Present findings add to existing knowledge by suggesting two personal dimensions of brand communication derived from the use case of CEO communication via Twitter. Furthermore, the study on hand is capable of contributing to science as personal dimensions prepare the ground to make the research quotient accessible for CEO reputation. We have identified novel dimensions that clearly distinguish personal brands from organizational brands. While our findings allow sufficient conceptualization, our study makes a first move toward theorizing the social capital of managers as an ether for reputation. Addressing the interaction with social media as exogenous information and communication tools is a research area of serious concern. Identified dimensions might not only help steering social media affordances more towards organizational needs. Platforms begin to integrate dashboard and analytics applications to meet the requirements of professional use. However, there is no such thing as an IT-based solution addressing reputational dimensions.

Our study comes with limitations as our findings rest upon a small sample of CEO and corporate brand pairs. Moreover, a snapshot of 8 weeks might not cover all facets of CEO reputation management performed on Twitter, especially for CEOs who author less frequently. We are aware that manual coding as part of qualitative research methods underlie the subjective assessment of all authors involved.

To strengthen our findings, more empirical research is required. We intent to extent the examination of this matter to a wider spectrum of social media platforms (e.g. LinkedIn, Facebook, Instagram) and industries to create a more heterogeneous data sample. Moreover, scholars researching this issue may consider falling back on a long-term data foundation and include critical phases of reputation management with increased volatility such as corporate crises. This includes cross-cultural examinations outside of the US-market. In order to build upon social media capacities that incorporate reputational dimensions, we suggest developing software prototypes that translate social media metrics to reputational scores, e.g. through automated sentiment analyses. This would allow organizations and CEOs to keep track of distinct levels of reputation and adjust communication strategies accordingly.

References

1. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media, *Bus. Horiz.*, 53(1), pp. 59–68 (2010)
2. Lin, Z., Goh, K.Y.: Social Media Brand Community and Consumer Purchase: a Cross-Brand Analysis, in *ECIS,* pp. 39-52 (2012)
3. Karjaluoto, H., Mäkinen, H., Järvinen, J.: A Firm's Activity in Social Media and Its Relationship with Corporate Reputation in *Bled eConference*, pp. 469–481 (2015)
4. Capriotti, P., Ruesja, L.: How CEOs use Twitter: A comparative analysis of Global and Latin American companies, *Int. J. Inf. Manage.*, 39(1), pp. 242–248 (2018)
5. Marx, J., Mirbabaie, M., Stieglitz, S.: You are now an Influencer! Measuring CEO Reputation in Social Media, in *ACIS*, pp. 1-10 (2018)
6. Sihi, D., Lawson, K.: Marketing Leaders and Social Media: Blending Personal and Professional Identities, *J. Mark. Theory Pract.*, 26(1–2), pp. 38–54 (2018)
7. Hayward, M. L. A., Rindova, V. P., Pollock, T.G.: Believing one's own press: The causes and consequences of CEO celebrity, *Strateg. Manag. J.*, 25(7), pp. 637–653 (2004)
8. Pamuksuz, U., Mourad, M.: "Analysis of Executive Reputation in Social Media Networks," *Proc. Thirty Seventh Int. Conf. Inf. Syst.*, pp. 1–9 (2014)
9. Floreddu, P.B., Cabiddu, F., Evaristo, R.: Inside your social media ring: How to optimize online corporate reputation, *Bus. Horiz.*, 57(6), pp. 737–745 (2014)
10. Dijkmans, C., Kerkhof, P., Beukeboom, C. J.: A stage to engage: Social media use and corporate reputation, *Tour. Manag.*, 47, pp. 58–67, (2015)
11. Gatzert, N.: The impact of corporate reputation and reputation damaging events on financial performance, *Eur. Manag. J.*, 33(6), pp. 485–499 (2015)
12. Benthaus, J.: Making the Right Impression for Corporate Reputation: Analyzing Impression Management of Financial Institutions in Social Media, In: *ECIS*, pp. 1–17 (2014)
13. Treem, J., Leonardi, P.: Social Media in Organizations: Exploring Affordances of Visibility, Editability, and Association, *Commun. Yearb.*, 36, pp. 143–189 (2012)
14. Scott, S., Walsham, G.: Reconceptualizing and Managing Reputation Risk in the Knowledge Economy: Toward Reputable Action, 16(3), pp. 308–322 (2005)
15. Aral, S., Dellarocas, C., Godes, D.: Introduction to the special issue - social media and business transformation: A framework for research, *Inf. Syst. Res.*, 24(1), pp. 3–13 (2013)
16. Fombrun, C. J.: *Reputation: Realizing Value from the Corporate Image*. Cambridge: MA: Harvard Business School Press, (1996)
17. Fombrun, C., Shanley, M.: What's in a name? Reputation building and corporate strategy, *Acad. Manag. J.*, 33(2), pp. 233–258 (1990)
18. Nguyen, N., Leblanc, G.: Corporate image and corporate reputation in customers' retention decisions in services, *J. Retail. Consum. Serv.*, 8(4), pp. 227–236 (2001)
19. Graham, M.E., Moore, J.: Consumers' willingness to pay for corporate reputation: the context of airline companies, *Corp. Reput. Rev.*, 10(3), pp. 189–200 (2007)
20. Smith, K.T., Smith, M., Wang, K.: Does brand management of corporate reputation translate into higher market value?, *J. Strateg. Mark.*, 18(3), pp. 201–221 (2010)
21. Mirbabaie, M., Stieglitz, S., Ruiz Eiro, M.: #IronyOff – Understanding the Usage of Irony on Twitter during a Corporate Crisis, In: *PACIS*, 1-16 (2017)
22. Schenk, C. B.; Sicker, D.C.: Finding event-specific influencers in dynamic social networks, In *IEEE*, pp. 501–504 (2011)

23. Park, D. J., Berger, B. K.: No The presentation of CEOs in the press, *J. Public Relations Res.*,16(1), pp. 93–125 (2004).
24. Ranft, A.L., Ferris, G.R., Zinko, R., Buckley, M.R.: The costs and benefits of CEO reputation, *Organ. Dyn.*, 35 (3), pp. 279–290 (2006)
25. Stieglitz, S., Mirbabaie, M., Potthoff, T.: Crisis Communication on Twitter during a Global Crisis of Volkswagen – The Case of "Dieselgate" In: *HICSS*, 1-14 (2018)
26. Tsai, W.H.S., Men, L.R.: Social CEOs: The effects of CEOs' communication styles and parasocial interaction on social networking sites, *New Media Soc.*, 19(11), pp. 1848–1867 (2017)
27. Karahanna, E., Xin Xu, S., Xu, Y., Zhang, A.: The Needs–Affordances–Features Perspective for the Use of Social Media, *MIS Q.*, 42(3), pp. 737–756 (2018)
28. Porter, M., Anderson, B., Nhotsavang, M.: Executive Twitter engagement and attitudes about media credibility, *J. Commun. Manag.*, 19(3), pp. 270–287 (2015)
29. Weng, P.S., Chen, W.Y.: Corporate reputation, CEO reputation, and corporate financial performance, *North Am. J. Econ. Financ.*, 39(1), pp. 223–240 (2017)
30. Fombrun, C.J., Gardberg, N.A., Sever, J.M.: The Reputation QuotientSM: A multi-stakeholder measure of corporate reputation, *J. Brand Manag.*,7(4), pp. 241–255 (2000)
31. Walsh, G., Beatty, S.E.: Customer-based corporate reputation of a service firm: scale development and validation, *J. Acad. Mark. Sci.*, 35 (1), pp. 127–143, (2007)
32. Kanto, D.S., de Run, E.C., bin M. Isa, A.H.: The Reputation Quotient as a Corporate Reputation Measurement in the Malaysian Banking Industry: A Confirmatory Factor Analysis, *Soc. Behav. Sci.*, 219, pp. 409–415 (2016)
33. Gardberg, N.A.: A Cross-Cultural Qualitative Analysis of Construct and Instrument Equivalence, *Corp. Reput. Rev.*, 9(1), pp. 39–61 (2006)
34. Wiedmann, K.-P., Fombrun, C. J., van Riel, C. B. M.: Reputationsanalyse mit dem Reputation Quotient," in *Handbuch Unternehmenskommunikation*, pp. 321–337 (2007)
35. Van Norel, P., Kommers, J., Van Hoof, J., Verhoeven, J.: Damaged corporate reputation: Can celebrity Tweets repair it?, *Comput. Human Behav.*, 36, pp. 308–315 (2014)
36. Deephouse D.L.: Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories, *J. Manage.*,26(6), pp. 1091–1112 (2000)
37. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems, In: *EJIS*, 22 (3), pp. 336–359, (2013)
38. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a Social Network or a News Media? Categories and Subject Descriptors, In: *WWW,* pp. 591-600, (2010)
39. Shariff, S. M., Zhang, X., Sanderson, M.: On the credibility perception of news on Twitter: Readers, topics and features, *Comput. Human Behav.*, 75, pp. 785–796, (2017)
40. Griffith, E.: Top 50 Global Technology Companies," In: *Fortune Magazin*, http://fortune.com/2015/06/13/fortune-500-tech/ (Accessed: 17.12.2018)
41. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in Twitter: The million follower fallacy, In: *AAAI*, 2, pp. 10–17 (2010)
42. Lee, E., Kim, J., Baik, D. K.: Method for Measuring Twitter Content Influence Subtitle as needed, In: *SEKE,* pp. 659–664 (2014)
43. Riquelme, F., González-Cantergiani, P.: Measuring user influence on Twitter: A survey, *Inf. Process. Manag.*, 52 (5), pp. 949–975 (2016)
44. Van Dijck, J., Poell, T.: Understanding Social Media Logic, In: *Media and Communication*, pp. 2–14 (2013)
45. Mayring, P.: Qualitative Inhaltsanalyse," In: *Handbuch Qualitative Forschung in der Psychologie*, Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 601–613, (2010)
46. Krippendorff, K.: Content analysis: An introduction to its methodology, 3rd edition. Thousand Oaks, CA: Sage, (2013)

A Multiorganisational Study of the Drivers and Barriers of Enterprise Collaboration Systems-Enabled Change

Clara S. Nitschke[1], Susan P. Williams[1], and Petra Schubert[1]

[1] University of Koblenz-Landau, Faculty of Computer Science, Koblenz, Germany
{cnitschke,williams,schubert}@uni-koblenz.de

Abstract. Enterprise Collaboration Systems (ECS) are emerging as the de facto technology platform for the digital workplace. This paper presents findings from an in-depth, multiorganisational study that examines the drivers and barriers of ECS-enabled change from two perspectives: i) the company initiating and driving the project and ii) key practitioners responsible for delivering the change. Data is collected from ECS using companies via a survey and face-to-face workshops, analysed using qualitative content analysis methods to identify categories of change and then synthesised to provide a rich classification and visualisation of the drivers, barriers, motivations and pain points (DBMP) to ECS-enabled change. This is followed by a discussion of the similarities and differences between drivers and barriers from both personal and company perspectives. The paper concludes by exploring the potential of the research and visualisation methods used in this work to provide the foundation for the longitudinal study of ECS-enabled change.

Keywords: Enterprise Collaboration System (ECS); Enterprise Social Software (ESS); ECS-Enabled Change; Drivers; Barriers.

1 Introduction

The transformation of work and new ways of working are essential for organisational success in the digital era [1]. Based on the success of social media platforms in everyday life [2], socially-enabled Enterprise Collaboration Systems (ECS) have emerged to extend traditional groupware (e.g. email, shared calendar) through the integration of social software features (e.g. wikis, blogs, social profiles) and provide large-scale integrated platforms to connect people, work practices, activities and structures [3], [4]. ECS have become the de facto IT platform at the heart of the digital workplace [5] and are generating significant interest for both researchers and practitioners [6].

Recent research has identified that initiatives for the introduction of ECS are being driven by company-specific objectives, including, for example, improved innovation management, better collaboration between employees, improved knowledge transfer and improved search for experts [7], [8]. However, individual organisations have differing drivers for introducing an ECS platform and are also experiencing a variety of barriers to system adoption and use, including for example cultural changes, reliance on other systems [9], [10] and poor formalisation of ECS communication protocols

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

[11]. Current research into the drivers and barriers of ECS projects typically draws from cross-sectional case studies conducted at a single point in time [10]. Both quantitative and qualitative research methods have been used to examine the adoption of ECS and enterprise social software (ESS) to identify and investigate ECS/ESS adoption drivers and barriers [12–19]. Nielsen and Razmerita [20] found that while ECS have the potential to support a variety of objectives, their adoption is affected by factors including individual (e.g. technical skills), organisational (e.g. management support) and technical factors (e.g. usability). By studying uncertainties as barriers for knowledge sharing in ECS, Trier et al. [18] place emphasis on the individual perceptions impacting and constituting barriers to ECS adoption, such as uncertainties regarding the purpose of the collaboration platform. While not making a distinction between ECS adoption barriers from a personal and a company perspective, Forstner and Nedbal [19] identified five problem areas: project management, technology, culture, top/middle management and employees. They also argue that organisations need to identify problems with the adoption of ESS at an early project stage and address them through project management. Similarly, Diehl et al. [2, p. 247] state that "cultural challenges can be anticipated and should be managed ex ante, not ad hoc." However, in line with the evolving and sociotechnical nature of ECS, the drivers and barriers organisations and their employees experience are emerging over time as they make sense of ECS and embed them into their digital workplace.

While current research provides insights into the importance of studying ECS adoption drivers and barriers, we see the need to establish appropriate means for capturing and investigating these drivers and barriers and building the foundation for studying the ways that they change over time. In their study on ECS adoption, Greeven and Williams [10] show that ECS adoption challenges and barriers are multifaceted and exemplify the complexity of ECS projects which organisations and the involved stakeholders are facing over time. In this paper, we build on our previous research and literature reviews on ECS adoption [8-10] and present the findings from an in-depth, multiorganisational empirical study that identifies drivers and barriers to ECS-enabled change from both the viewpoint of the company initiating the project and key practitioners involved in the everyday implementation and management of the project. Our goal is to extend existing work and lay the foundation for a long-term, longitudinal view of ECS-enabled change by following the introduction and use of ECS platforms over time.

2 Research Design and Data Analysis

This study is part of IndustryConnect, a long-term university-industry research program in the field of collaboration technologies and the design of the digital workplace [21]. IndustryConnect brings a team of researchers from a German University together with key practitioners from 31 German/Swiss companies. The organisations participating in the study are leaders in the introduction and use of ECS and have implemented IBM Connections, currently the largest, most integrated ECS platform [5]; all are committed to participating in interactive research to examine ECS-enabled change in their organisations. The IndustryConnect member companies are typically medium- and large-sized (1,000-300,000 employees) representing a range of industry sectors including: manufacturing, transport/logistics, retailing, government services and financial services. The key practitioners are all employees with responsibility for the introduction and use of the ECS in their organisation and represent a range of professional backgrounds including: information technology, information/ knowledge management, internal communications, organisational processes and business development.

The aim of this study is to identify and understand the drivers (D), barriers (B), motivations (M) and pain points (P) to ECS-enabled change and the adoption and ongoing development of ECS in organisations. Our objective is to examine ECS-enabled change from two perspectives:

i) Company perspective. The focus is on the company context and the broader reasons for initiating and driving the ECS project. From the company perspective we identify: *drivers (D)*, reasons why the company originally initiated the ECS project (e.g. expressed in strategy documents or strategic plans); and *barriers* (B), challenges/problems encountered that are constraining the ECS project and making the achievement of the company's ECS-enabled change objectives more difficult, and

ii) Personal perspective. The focus is on the everyday motivations and challenges of ECS-enabled change from the perspective of the key practitioners responsible for delivering the change. This perspective is shaped by individual experiences and the everyday activity in the organisation [22]. From the personal perspective we identify: *motivations (M),* reasons the ECS project is being supported or pushed forward by the individual practitioner and *pain points (P),* current challenges, problems, issues in the ECS project that the individual practitioner experiences in his/her daily work on the project and making the achievement of ECS-enabled change objectives more difficult.

The research study design is structured into four phases as shown in Figure 1 and discussed below.

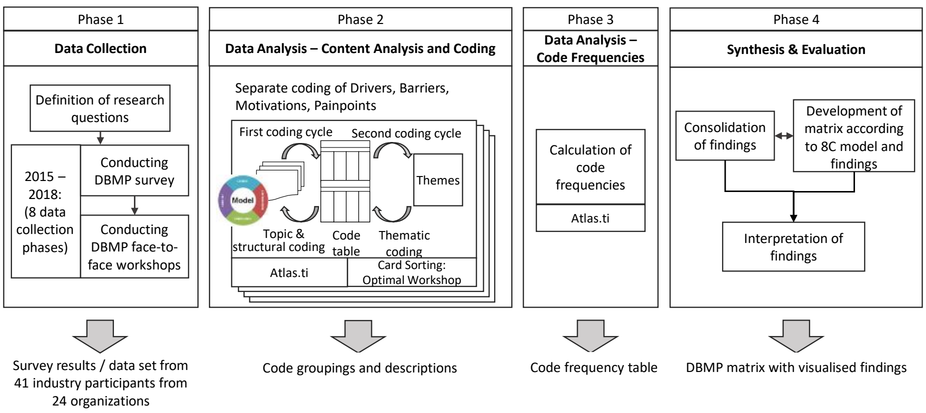


Figure 1. Research design phases

Phase 1: Data Collection. Primary data about drivers, barriers, motivations and pain points (DBMP) related to ECS introduction and use were collected between 2015 and 2018 from 41 practitioners representing 24 companies from the IndustryConnect program. Data was first collected via an online questionnaire comprised of four questions addressing both the company perspective ("What are the original *drivers* for your company initiating the ECS project?", "What are the key *barriers* encountered by your company constraining the ECS project?") and the personal perspective ("What are your current personal *motivations* for supporting the ECS project or pushing it forward?", "What are your personal *pain points* in the ECS project you experience in your daily work on the project?") of key practitioners involved in the respective ECS-enabled change projects. After completing the questionnaire respondents presented their DBMP in one of a series of face-to-face workshops. The objective of the workshop presentation is for each respondent to elaborate on their answers, clarify meanings and for the researchers to gather further examples. The presentations and discussions were digitally recorded and transcribed for analysis.

Phase 2: Data Analysis – Content Analysis and Coding. The drivers, barriers, motivations and pain points (DBMP) identified by each respondent were listed and coded. Using content analysis methods and following Saldaña [23] the data was coded through two iterative coding cycles using the qualitative data analysis tool ATLAS.ti. In the first coding cycle, topic and structural coding was applied to provide four basic code tables for the DBMP. Guided by the elements of the 8C model [24], the codes in the individual tables were then structured into two categories: i) functional DBMP codes relating to the ECS platform, its functionality and performance and ii) business DBMP codes relating to the organisation and its corporate objectives.

In the second coding cycle the card sorting method for thematic coding was applied [25]. To ensure research quality and reliability three researchers worked independently on the coding and reviewed the codes in joint workshops to clarify the emerging DBMP categories and to achieve a high level of intercoder reliability. Following the coding process, tables containing DBMP code groupings and code descriptions were created.

Phase 3: Data Analysis – Code Frequencies. Code frequencies, that is the total number of occurrences of each code, were calculated in ATLAS.ti and the code frequency table was created. The code frequency table provides an additional means for examining similarities and differences between drivers, barriers, motivations and pain points for different practitioners and companies.

Phase 4: Synthesis and Evaluation. In the final phase of the study the findings were synthesised and consolidated. The DBMP matrix (Figure 2) was developed to display the consolidated findings in a visual form. These consolidated results were presented to the study participants in a review workshop. The DBMP matrix itself was also evaluated as a method for visualising and consolidating the drivers, barriers, motivations and pain points data and as a method for presenting the data to the participants. The feedback from the participants was positive and the matrix provoked constructive discussions between the researchers and the study respondents.

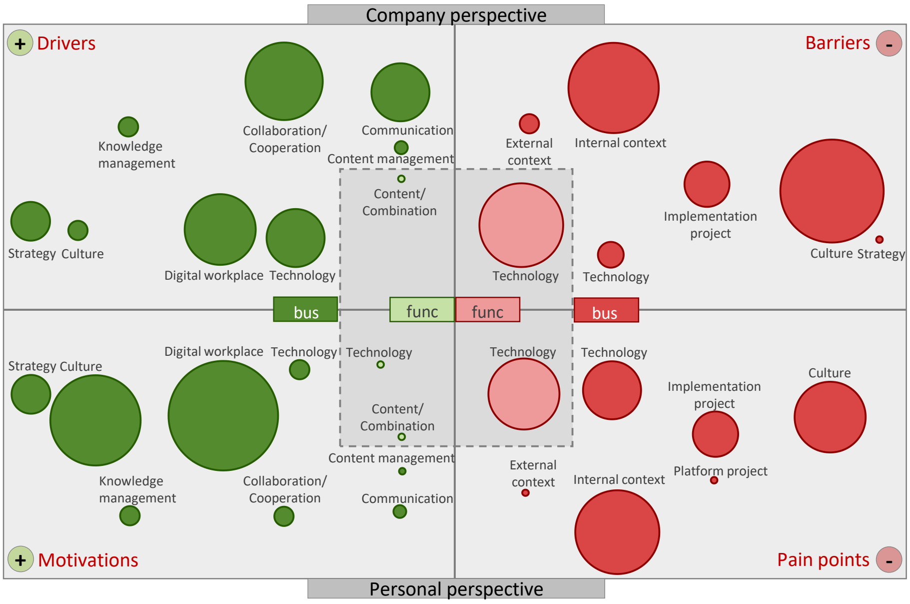


Figure 2. Drivers, barriers, motivations and pain points of ECS-enabled change

3 Findings and Discussion

The derived DBMP categories were visualised in a 2x2 matrix (Figure 2). The top two quadrants represent the company view, the bottom two quadrants the personal view; the quadrants on the left show the issues driving the project forward (*drivers* and *motivations*, green) and the quadrants on the right show the issues constraining the project (*barriers* and *pain points*, red). In total, 108 different DBMP codes were identified in the final coding iteration. The identified DBMP codes and categories were separated into *functional* (light green / light red at the centre of Figure 2) and *business* (dark green / dark red in Figure 2). The size of the bubbles represents the sum of the frequencies of the codes within a category. To enhance readability, bubbles of similar categories are placed next to each other (e.g. content management and communication, that can be found in the 8C model [24], or internal and external context). Furthermore, the matrix lines work roughly as mirror axes to improve comparability between the single quadrants and between functional and business DBMP within one quadrant.

Table 1. Drivers/Motivations: category descriptions and frequencies

Category		*Description*	*D*	*M*
Collaboration / Cooperation	bus	Relate to new or changing collaboration and cooperation practices and processes for business improvement	12	3
Communication	bus	Relate to new or changing communication practices and processes for business improvement.	9	2
Content Management	bus	Relate to new or changing content management practices and processes for business improvement.	2	1
Knowledge Management	bus	Relate to new or changing knowledge management practices and processes for business improvement.	3	3
Digital Workplace	bus	Relate to an integrated collaboration platform and what it is expected to enable regarding the support of organisational members and their work.	11	17
Technology	bus	Relate to the improved management and usage of IT within the corporate context and desired changes to business IT (architecture) solutions and concepts.	9	3
Strategy	bus	Relate to corporate strategy/ strategic objectives.	6	6
Culture	bus	Relate to the corporate culture, where cultural changes need to be enabled through the introduction of new technology and work practices and/or are needed for the successful embedding of new technology and work practices into the work environment.	3	14
Content / Combination	func	Relate to content creation, use, and management functionally supported by the collaboration platform.	1	1
Technology	func	Relate to expected or perceived positive characteristics and affordances of the collaboration platform itself and its functionality, e.g. ease of use or integration capability.	x	1

A similar set of categories emerged for the drivers (D) and motivations (M), and for the barriers (B) and pain points (P). Table 1 and Table 2 show the descriptions of the business (bus) and functional (func) categories identified for the four quadrants (DBMP) and the associated frequencies of the code categories. A cross is used in the tables when a category is not present in the respective DBMP quadrant. Each category description is comprehensive to cover a wide range of different DBMP codes. For example, the barriers and pain points category *implementation project* covers ten codes: *lack of systematics in the introduction*, *neglect of the head quarter*, *missing business link*, *poor transparency*, *low awareness level*, *poor training*, *missing binding rules*, *different target visions*, *missing use cases*, and *poor project management*.

Table 2. Barriers/Pain points: category descriptions and frequencies

Category		*Description*	*B*	*P*
External context	bus	Relate to actors and influences from outside the company that are restricting, constraining or otherwise negatively impacting the ECS project and its development (e.g. laws, regulations, customers, vendors, market developments).	3	1
Internal context	bus	Relate to actors and influences from inside the company that are restricting, constraining or otherwise negatively impacting the ECS project and its development (e.g. poor management support of the ECS project, work council, company structures, missing resources).	14	13
Implementation project	bus	Relate to the ECS implementation project itself and corresponding implementation strategy design decisions.	7	7
Platform management	bus	Relate to the ways the collaboration platform is managed and the consequences thereof.	x	1
Technology	bus	Relate to the management and usage of IT within the corporate context and unsatisfactory business IT (architecture) solutions and concepts.	4	9
Strategy	bus	Relate to the embedding/alignment of the collaboration platform into/with the corporate strategy.	1	x
Culture	bus	Relate to corporate culture, where cultural changes are needed for the successful embedding of new technology and work practices into the work environment.	16	11
Technology	func	Relate to perceived deficiencies in the collaboration platform itself and its functionality and usability, e.g. security, missing functionality, external access.	13	11

As can be taken from Table 1 and Table 2, there are differences between the personal and the company view in terms of the total number of mentions within a category. In the following sections we describe and discuss each of the four quadrants with illustrative examples of associated codes.

3.1 Drivers

Eight business driver categories and one functional driver category were identified as shown in Table 1. Each driver category contains a set of driver codes, for example, the driver category *collaboration/cooperation* contains driver codes such as: *increase of inter-site collaboration* or *support of international collaboration*.

The frequency of occurrence of the single driver codes indicates a stronger focus on business drivers and less on functional-oriented drivers for the introduction of an ECS. Since the drivers represent the company perspective, it is not surprising that there are more business-oriented driver codes and categories reflecting overall corporate objectives and visions, such as globalisation, business performance, productivity improvement, or support of organisational changes [26], [27]. The analysis reveals that ECS drivers may be i) opportunity-driven (e.g. support of international communication, improvement of cross-divisional collaboration, sharing of information) or ii) problem-driven (e.g. usage of external social media applications, communication hierarchy, reduction of e-mails).

Table 3. Driver examples

Quote	*Driver code*
"A global collaboration tool for all employees to bring colleagues from around the world closer together, in a flexible, forward-looking network culture" (Service Owner (IT), Automotive Parts Manufacturer 05)	D: bus: collaboration/cooperation: global collaboration tool; D: bus: digital workplace: networking
"The communication hierarchy did not fit; turn hierarchies upside down" (Head of IT, Security/Infrastructure, Clothing Manufacturer/Retail 02)	D: bus: communication: communication hierarchy
"Support cross-border collaboration" (Head of Communications/Knowledge, Consumer Electronics Manufacturer 02)	D: bus: collaboration/cooperation: support of international collaboration

In general, the main focus of drivers is on business achievements and improvements. In particular, organisations are striving to enable and improve collaboration and communication between employees as well as becoming a digital workplace and having the right technology in place to achieve this. The collaboration platform itself and its functionality and performance are of less concern as drivers of change. Table 3 shows examples of typical drivers expressed by the study participants.

3.2 Barriers

Six categories of business barrier and one functional barrier category were identified (Table 2). As with the drivers, each barrier category contains a set of barrier codes. For example, the barrier category *implementation project* contains barrier codes such as *missing binding rules* and *poor transparency*. The barrier code frequencies reveal that the biggest groups of responses relate to the business categories *internal context* and *culture* and to the functional category *technology*. Internal context largely refers to people who constrain the project (e.g. managers not supporting the project or the works council demanding specific requirements regarding personally identifiable information being captured in the ECS) and is seen as particularly problematic. Culturally, change management and the development of a new mindset is perceived as challenging. Functional barriers pointing to weaknesses in the collaboration platform (e.g. in terms of usability) and functional limitations (e.g. insufficient analytics measurement capabilities) were identified as major barriers to the ECS-enabled change projects.

Table 4. Barrier examples

Quote	*Barrier code*
"Knowledge is seen as property from the employees' point of view. Knowledge is not shared because this [new] attitude is not rooted yet. Also, this has something to do with the notion of performance. The employee thinks ‚If I advance something, I will be in the favour of my boss'" (Specialist, Collaboration, Steel Manufacturer 01)	B: bus: culture: corporate culture that makes collaboration difficult
"Financial and human resources" (Project Manager, Information Architecture, Logistics Services 01*)*	B: bus: internal context: missing resources
"The ECS is far too extensive, security and privacy requirements destroy the usability of the product. The app is not at all user-friendly." (Head of Corporate Portals, Air Transportation 01)	B: func: technology: system complexity; B: func: technology: usability

Barriers relate primarily to business issues, however, in contrast to the drivers, functional barriers are also of importance. The identified codes and categories show that some barriers only became visible in ECS use; they were unanticipated and organisations only recognized them through using the platform. For example, a missing killer app or system complexity that only becomes visible when the system is in use (Table 4). Functional deficiencies of the collaboration platform make achieving expected ECS benefits and outcomes, e.g. improved communication, more difficult.

3.3 Motivations

The motivation categories identified in the data analysis are similar to the driver categories, however they represent the personal view of the key practitioners and include different objectives. Eight business motivation categories and one functional motivation category were identified (Table 1). As with the drivers and barriers, each category contains a set of motivation codes. For example, for the motivation category *digital workplace* the motivation codes *creating transparency* and *establishing modern ways of working* were identified. These motivations focus on opportunities, such as designing the workplace of the future, establishing new ways of working, or promoting changes in the corporate culture.

Table 5. Motivation examples

Quote	*Motivation code*
"The opportunity to fundamentally make the daily work of 70.000 employees easier." (Senior Manager, HR, Logistics Services 02)	M: bus: digital workplace: make work easier
"I want to help shape change." (Senior Manager, Social Collaboration, Automotive Parts Manufacturer 02)	M: bus: culture: promote change/ change in corporate culture
"Future Work is my mission." (Manager Digital Transformation and Change, Automotive Parts Manufacturer 05)	M: bus: digital workplace: workplace of the future

The personal motivations in ECS projects are mainly in the business categories *culture* and *digital workplace*. The frequencies of the motivation codes and categories revealed the individual practitioners desire to be part of organisational change and the shaping and transforming of the company's culture and digital workplace. Functional motivations are seen as being of minor importance. The collaboration platform itself and its performance is of less importance; instead shaping the digital transformation is paramount. Table 5 provides examples of motivations where the active part and role of the individual participants in the ECS-enabled change projects becomes clear.

3.4 Pain Points

Five business pain point categories and one functional pain point category were identified (Table 2) where each category includes a set of pain point codes. For example, for the pain point category *culture* pain point codes of *change management* or *poor support for new mindset / ways of working* were identified. In line with the set of categories and the code frequencies, pain point categories (personal view) and barrier categories (company view) show highest similarity. As with the barrier categories, the categories *internal context* (business), *culture* (business), and *technology* (functional) are the most frequently cited categories, followed by *technology* (business). From the perspective of the key practitioners involved, the ECS project is constrained from inside the company due to missing resources and resistance from, for example, the works council as well as low level of digital competence. Culturally, acceptance and the lack of support for a new mindset and work practices is perceived as most challenging. Worth noting are the technology aspects that are being negatively perceived. This applies to both the corresponding business category, e.g. through multiple possible competing systems used by the organisation, and the corresponding functional category with issues of technology usability and system integration.

Table 6. Pain point examples

Quote	*Pain point code*
"Too many different tools and no single point of entry" (IT Manager, Collaboration/Knowledge, Technology Inspection Service)	P: bus: technology: variety of systems P: func: technology: missing single sign-on
"Binding rules on collaborative work [...] regulate a particularly strong group of Confluence users." (Project Manager Intranet, Retail Grocery 01)	P: bus: implementation project: missing binding rules; P: bus: technology: competing system
"Works council." (Project Manager, Automotive Parts Manufacturer 04)	P: bus: internal context: slow movement due to the works council
"System acceptance and the competition with Microsoft products" (Internal Consultant, Air Transportation 01)	P: bus: culture: acceptance; P: bus: technology: competing system
"Deficits in IBM Metrics: Development of analytics methods without existing documentation" (Specialist, Collaboration/ Statistics, Automotive Parts Manufacturer 02)	P: func: technology: deficits in metrics capability

As with the barriers, functional technology-related pain points become visible through ECS use. For example, practitioners noticed the significance of pain points such as a missing single sign-on feature when the ECS is being used alongside multiple other business software systems. The participants' responses in Table 6 show examples of individually perceived pain points in the ECS projects that make shaping the new workplace (which was one of the key motivations) more difficult.

3.5 Interpretation of Drivers, Barriers, Motivations and Pain Points

The visualisation of the DBMP (Figure 2) and the underlying data reveal both similarities and differences in the collected drivers, barriers, motivations and pain points.

While drivers and motivations for bringing the ECS projects forward are largely business-oriented, barriers and pain points clearly deal with functional issues related to the collaboration platform itself, including deficiencies in data analytics capabilities, external access, integration or usability.

From the study findings we also see a link between the position and role of the individual participants in the ECS project and the nature (functional or business) of the perceived pain points. As can be drawn from the example pain points (Table 6) the business-oriented pain points were largely identified by project managers and internal consultants, while the functional-oriented pain points were identified by IT managers and technology specialists. Overall, however, business-oriented DBMP are most prevalent. From a company perspective, the business-oriented drivers are about enabling and/or improving things, such as collaboration or the digital workplace, and place emphasis on ECS project outcomes. In contrast, business-oriented motivations showing the personal perspective focus on the path towards these outcomes, where having an active part in the transformation, particularly of the organisational culture and digital workplace, is key. Such obvious differences cannot be directly identified from the two quadrants on the right side of the matrix but are revealed in the coding tables. Both, business-oriented barriers and pain points specifically represent challenges from the internal context, for example through groups of people who obstruct or constrain the ECS project and the changing of culture. In addition, the business-oriented category *implementation project* has the same frequency for barriers and pain points. One reason for this is that that pain points personally perceived by the people responsible for the ECS project (e.g. project leaders and managers) are partly based on the ECS barriers prevailing inside the company, as these must be addressed by them in order to achieve ECS acceptance and satisfy their individual ECS project motivations.

Furthermore, there are some categories, e.g. *culture* or *technology* that can both be perceived as driving and constraining the ECS project. For example, while organisations and key actors in the ECS projects aim to bring about changes in the corporate culture and collaboration mindset by introducing a collaboration platform, the current cultural mindset and attitudes might also impede accepting and adopting new technologies and work practices.

4 Concluding Remarks

In this paper, we identify and analyse the drivers and barriers to the adoption of ECS. We conducted a survey with multiple ECS user organisations from different industries. Since ECS are largely being shaped by individual key practitioners in the organisations, who have different backgrounds, experiences with and expectations of collaboration systems, the *drivers* and *barriers* were studied from both a company perspective and a personal perspective. In order to differentiate between these views, we named the drivers and barriers from a personal perspective *motivations* and *pain points*. The drivers, barriers, motivations and pain points (DBMP) were collected, coded and their frequencies identified. This data is consolidated and visualized in the DBMP matrix, displaying the diverse DBMP categories and their relevance for ECS user organisations. This work extends previous work, for example [13], [16], [19] by representing multiple perspectives (company and personal) and providing a deeper and more nuanced classification of drivers and barriers.

While the current study does not yet consider the ways that DBMP change over the life of ECS projects it has delivered a clear set of DBMP anchor measures [28] that can be traced over time. The research approach and DBMP matrix provide a method for collecting and visualising DBMP and are now being used to capture DBMP from the same group of ECS using companies at regular points in time; enabling us to examine how they change (or not) over time. To achieve this requires a more animated visualisation of the DBMP matrix that incorporates the dimension of time.

From this study and further interviews and workshops with the current participants we have further identified that companies with similar drivers and motivations are designing the digital workplace differently [5] and are dealing with the same barriers and pain points in different ways. For example, the barrier and pain point of the works council being perceived as impeding the ECS project progress (by having specific requirements regarding personally identifiable information in the system or the inclusion of certain employee groups), is being addressed differently in different organisations. While some organisations involve and incorporate the works council in the ECS project, others attempt to exclude them or redirect attention away from seemingly problematic ECS functionality and procedures.

Building on this DBMP study, work is now underway to study how DBMP evolve and change over time as part of an organisation's digital transformation efforts. Our goal is to identify the organisational competencies and capabilities required to successfully achieve and manage ECS-enabled change. In particular, we are identifying key enablers and constraints to specific ECS outcomes in order to gain insights into how the ECS transformation process is being shaped. The achievement of expected and desired outcomes, e.g. faster innovation or removal of knowledge silos, requires the development of competencies and capabilities allowing for the management of ECS-enabled change and the successful embedding of ECS into the digital workplace.

Throughout the digital transformation process companies are encountering both expected and unanticipated enablers and constraints to desired ECS outcomes. For example, companies have benefitted from positive use cases that make the ECS more visible and encourage employees to be more accepting of social software. On the other hand, they are constrained by, for example, conflicting stakeholder interests and responding to new regulatory requirements. Existing research on IS capabilities has identified that capabilities are developed through action and interaction with technology and the embedding of emerging skills and competencies within the organisation [29–31]. Additionally, the emergence of capabilities requires a process of reflection and learning embedded in the specifics of the organisational context [32]. In order to successfully build a digital transformation capability, we see the need for organisations to i) reflect on and learn from the digital transformation process while identifying and developing the relevant competencies and resources, and to ii) anticipate future changes shaping the digital workplace while building the knowledge, skills and resources for enabling digital change. In this way and in contrast to prior research on the ECS introduction and adoption [15], [17–20], this research is studying ECS as evolving and sociotechnical systems and follows Dourish [33] by viewing the change context as being dynamically designed through ongoing interactions with the collaboration system.

5 Acknowledgements

Supported by the Deutsche Forschungsgemeinschaft (DFG) Research Grant No. 285717372.

References

1. Dery, K., Sebastian, I.M., van der Meulen, N.: The Digital Workplace is Key to Digital Innovation. MIS Q. Exec. 16, 135–152 (2017)
2. Diehl, R., Kuettner, T., Schubert, P.: Introduction of enterprise collaboration systems: In-depth studies show that laissez-faire does not work. In: 26th Bled eConference. eInnovations: Challenges and Impacts for Individuals, Organizations and Society, pp. 236–250. Bled, Slovenia (2013)
3. Star, S.L., Ruhleder, K.: Steps toward an ecology of infrastructure: Design and access for large information spaces. Inf. Syst. Res. 7, 113–134 (1996)
4. Herzog, C., Richter, A.: Use Cases as a Means to Support the Appropriation of Enterprise Social Software. In: 49th Hawaii International Conference on System Sciences (HICSS 2016), pp. 4071–4080. Koloa, HI, USA (2016)
5. Williams, S.P., Schubert P.: Designs for the Digital Workplace. In: Proceedings of the CENTERIS - Conference on ENTERprise Information Systems. Lisbon, Portugal (2018)
6. Riemer, K., Stieglitz, S., Meske, C.: From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks. Bus. Inf. Syst. Eng. 57, 197–212 (2015)
7. Tierney, M.L., Drury, J.: Continuously improving innovation management through enterprise social media. J. Soc. Media Org. 1, 1–16 (2013)
8. Williams, S.P., Schubert, P.: Social Business Readiness Survey 2014. Koblenz: CEIR Research Report, No. 01/2015, University of Koblenz-Landau, Germany (2015)
9. Williams, S.P., Hausmann, V., Schubert, P., Hardy, C.A.: Enterprise 2.0 research: Meeting the challenges of practice. In: Proceedings of the 26th Bled Conference. Bled, Slovenia (2013)
10. Greeven, C.S., Williams, S.P.: Enterprise collaboration systems: Addressing adoption challenges and the shaping of sociotechnical systems. Int. J. Inf. Syst. Proj. Manag. 5, 5–23 (2017)
11. Osch, W. van, Steinfield, C.W., Balogh, B.A.: Enterprise social media: Challenges and opportunities for organizational communication and collaboration. In: 2015 48th Hawaii International Conference on System Sciences, pp. 763–722. Kauai, HI, USA (2015)
12. Kügler, M., Dittes, S., Smolnik, S., Richter, A.: Connect Me! Antecedents and Impact of Social Connectedness in Enterprise Social Software. Bus. Inf. Syst. Eng. 57, 181–196 (2015)
13. Meske, C., Stieglitz, S.: Adoption and Use of Social Media in Small and Medium-Sized Enterprises. 151, 61–75 (2013)
14. Choudrie, J., Zamani, E.D.: Understanding individual user resistance and workarounds of enterprise social networks: the case of Service Ltd. J. Inf. Technol. 31, 130–151 (2016)
15. DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., Muller, M.: Motivations for Social Networking at Work. In: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, pp. 711–720. ACM, San Diego, CA, USA (2008)
16. Richter, A., Stocker, A., Müller, S., Avram, G.: Knowledge management goals revisited: A cross-sectional analysis of social software adoption in corporate environments. Vine. 43, 132–148 (2013)
17. Antonius, N., Xu, J., Gao, X.: Factors influencing the adoption of Enterprise Social Software in Australia. Knowledge-Based Syst. 73, 32–43 (2015)
18. Trier, M., Fung, M., Hansen, A., Capili, A.: Uncertainties as Barriers for Knowledge Sharing with Enterprise Social Media. In: ECIS 2017 Proceedings, pp. 1619–1630. Guimarães, Portugal (2017)
19. Forstner, A., Nedbal, D.: A problem-centered analysis of enterprise social software projects. Procedia Comput. Sci. 121, 389–397 (2017)

20. Nielsen, P., Razmerita, L.: Motivation and knowledge sharing through social media within Danish organizations. In: Bergvall-Kåreborn, B. and Nielsen, P.A. (eds.) Creating Value for All Through IT. TDIT 2014. IFIP Advances in Information and Communication Technology, vol 429, pp. 197–213. Springer, Berlin, Heidelberg (2014)
21. Williams, S.P., Schubert, P.: Connecting industry: Building and sustaining a practice-based research community. In: Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS), pp. 5400–5409 (2017)
22. Mead, G.H.: Mind, self and society. University of Chicago Press, Chicago (1934).
23. Saldaña, J.: The coding manual for qualitative researchers. SAGE Publications, Los Angeles, CA (2009)
24. Williams, S.P.: Das 8C-Modell für kollaborative Technologien. In: Schubert, P. and Koch, M. (eds.) Wettbewerbsfaktor Business Software, pp. 11–21. Hanser, München (2011)
25. Nurmuliani, N., Zowghi, D., Williams, S.P.: Using Card Sorting Technique to Classify Requirements Change. In: Proceedings of the 12th IEEE International Requirements Engineering Conference (RE'04), pp. 240–248. IEEE, Kyoto, Japan (2004)
26. Murphy, K.E., Simon, S.J.: Intangible benefits valuation in ERP projects. Inf. Syst. J. 12, 301–320 (2002)
27. O'Leary, D.E.: Enterprise resource planning (ERP) systems: An empirical analysis of benefits. J. Emerg. Technol. Account. 1, 63–72 (2004)
28. Keen, P.G.W.: Shaping the future: business design through information technology. Harvard Business School Press, Boston, MA, USA (1991)
29. Penichet, V.M.R., Marin, I., Gallud, J.A., Lozano, M.D., Tesoriero, R.: A classification method for CSCW systems. Electron. Notes Theor. Comput. Sci. 168, 237–247 (2007)
30. Peppard, J., Ward, J.: Beyond strategic information systems: Towards an IS capability. J. Strateg. Inf. Syst. 13, 167–194 (2004)
31. Peppard, J., Ward, J., Daniel, E.: Managing the Realization of Business Benefits from IT Investments. MIS Q. Exec. 6, 1–12 (2007)
32. Andreu, R., Ciborra, C.: Organisational learning and core capabilities development: The role of IT. J. Strateg. Inf. Syst. 5, 111–127 (1996)
33. Dourish, P.: What we talk about when we talk about context. Pers. Ubiquitous Comput. 8, 19–30 (2004)

The More the Merrier? The Effect of Size of Core Team Subgroups on Success of Open Source Projects

Leonard Przybilla [1], Maximilian Rahn[1], Manuel Wiesche[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Chair for Information Systems, Garching, Germany
{leonard.przybilla,wiesche,krcmar}@in.tum.de
ga89rux@mytum.de

Abstract. Open source software (OSS) has become an important organizational form of building software. Given the desire to understand drivers of OSS project success and the known importance of social structure for team functioning, we investigate the effects of the relative size of contribution-based subgroups on community size of OSS projects. Drawing on extant research on OSS and faultline-based subgrouping, we investigate the relation with project community size of the relative size of subgroups based on reputation, issue focus, contribution extent and contribution persistence. While in several instances non-significant, results suggest a differential relation in which a large share of core members with high reputation, issue focus and persistent contributions positively relate to community size, whereas a large share of extensively contributing members in the core team is negatively related. Our findings are of value to research and practice by furthering the understanding of work in OSS projects.

Keywords: Open Source Software, Subgroups, Community Size, Team Governance.

1 Introduction

Open source and related concepts such as libre or free software development (in the following summarized as Open Source Software or OSS) have gained much traction in the beginning of the century [1] and continue to garner research attention recently [2].

Since most of the members of OSS projects contribute during their spare time and without monetary remuneration, the questions what motivates people to join, to contribute over longer periods of time and how such informal communities are managed have emerged as topics of research. Such issues are all the more relevant since despite overall success of OSS, a large majority of projects is defunct and not maintained [3]–leading to the issue how success in projects can be propelled.

OSS development is characterized as a virtual, distributed form of teamwork in which theoretically anyone can contribute [4]. This implies that developers are likely to differ on a number of attributes. OSS team members' motivations have been found

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

to be manifold–ranging from personal gain such as programming knowledge [1] or reputation [4] to philanthropic intentions [1]. Moreover, OSS team members embrace a common set of specific values and attitudes, which directly relate to work practices [5]. As may be expected from the clique-like description of OSS team interactions, the onboarding process of new members can be riddled with challenges [6].

The notion of positive effects of diversity in members on problem-solving [7] is only to some extent replicated in OSS [8], leading to the questions when and how diversity is conducive to outcomes. Based on characteristics shared by some members of a team and thus separating them from others, diversity can lead to so-called faultlines [7], which in turn may lead to perceivable subgroups [9]. Contingent on the specific reason for formation and configuration of faultlines and subgroups, the direction, i.e. enhancing or harmful, and strength of effects may differ [10, 11].

OSS teams as inherently open entities with a diverse set of members harbor much potential for faultlines and subgroups. Despite the critical importance of joint work in OSS, subgrouping and the resulting configuration as influential phenomena in general group research have, to the best of our knowledge, not received any research attention. We take a first step to addressing this void by investigating the following research question: Does the configuration of contribution-based subgroups in OSS teams relate to success as indicated by community size?

We first provide background information on extant research on OSS as well as faultlines and subgrouping before discussing specific implications in the context of OSS. We then introduce our hypotheses and the method used before reporting results of analysis and discussing implications. Lastly, we provide concluding remarks.

2 Background

In the following, we will briefly introduce extant research on OSS development and subgrouping before discussing the implications of subgroups in the context of OSS.

2.1 Work in Open Source Software Development

The success of OSS is astounding given its organizational challenges. Howison and Crowston (2014) report that organizing OSS is especially difficult for at least three reasons: Challenges presented by distributed work are exacerbated by relying on volunteers, which renders traditional incentive mechanisms ineffective. In addition, the work undertaken in OSS is complex with the associated difficulties [12]. Against the backdrop of previous research on the personality of developers [13], these assertions give rise to the questions which mechanisms help achieve valuable outcomes and why developers join and continue participating in OSS projects in the first place.

The motivations to join OSS projects are manifold. Given its characterization by voluntary contributions [1] and the ensuing absence of monetary remuneration, other causes such as personal motivations prevail. As private benefits, personal need for the developed functions, fun derived from working on the task and learning are key [1].

Membership in the community, the ability to gain reputation, and the possibility to receive job offers are also recurring themes [4]. The strong sense of community is mirrored in OSS participants sharing a common set of beliefs and values [5].

Assuming members assemble in a project, the issue of how work is organized arises. While OSS can be compared to several paradigms of work organization, unique differences are highlighted. Due to its inherently distributed nature, OSS can arguably be related to such teams, albeit results on governance may not be directly transferrable [12]. The voluntary and thus indeterminate nature is mirrored in elements that OSS development shares with agile projects [14]. More testament to the specific type of work accomplished in OSS is given by structural investigations. Typically, a relatively small core of developers contributes the majority of work, which is augmented by the smaller contributions of peripheral members [15]. Considering team composition, strong network ties of developers have been observed to bolster success [16]. For embeddedness of developers and projects, differential effects on success have been observed [17]. Moreover, the proficiency of projects at either developing new features or improving upon existing code has been observed to depend on the structure of collaboration [2].

The presence of a strong sense of community coupled with findings that a core of developers contributes differently than a periphery of developers gives rise to the question on how the configuration of the core relates to success in the larger community. To determine such possible effects, we propose to draw on faultline and subgroup theory.

2.2 Faultlines and Subgrouping

Diversity, i.e. differences in team members regarding attributes such as gender or functional background is found to be conducive to performance in teams by enabling the integration of diverse viewpoints [7]. In OSS, diversity of members has been found to improve some but not all outcomes [8].

Effects of diversity can be explained by so-called faultlines: Latent divisions among members based on characteristics shared by only some [7]. If perceived by members, faultlines are activated and lead to subgroups [18], i.e. several smaller entities within the overarching work teams [11]. For the purpose of this research, the term "subgroup" refers to activated faultlines and is rooted in faultline theory–notwithstanding its use in other contexts.

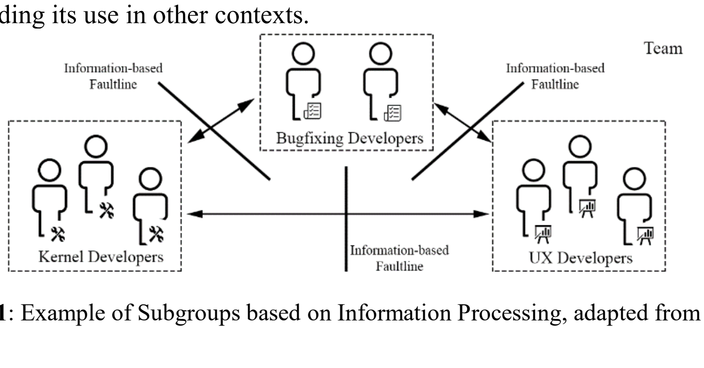


Figure 1: Example of Subgroups based on Information Processing, adapted from [19]

Faultlines and to a larger extent active subgroups have been found to affect team outcomes [9]. Recently, it has been proposed that the reason for subgroup formation may lead to different types of subgroups with different internal processes and thus different effects on group outcomes [11]. Identity-based subgroups due to e.g. differences in age are expected to trigger mostly negative processes, resource-based subgroups due to e.g. status differences harbor the potential for conflict but can boost efficiency, and information-based subgroups due to e.g. different expertise can engender team effectiveness by supporting information processing across groups. Given the need to coordinate knowledge in software development [20], the implications of information-based subgroups could be especially positive. Figure 1 exemplifies the emergence of subgroups based on information-triggered faultlines. The number of subgroups and their balance in terms of membership size, i.e. equally split versus imbalanced subgroups, also influence subgroup effects with e.g. an imbalanced configuration of geographically dispersed members leading to negative effects [10, 11]. Empirically, a complex interaction of subgroup formation, configuration and team outcomes has been observed [21]. In particular, software engineering practices may change subgrouping and its effects [19, 22].

2.3 Faultlines and Subgroups in Open Source Software

Considering the importance of commonly held values and community in OSS [5] and the observed effects of subgrouping raise the question whether harmful or positive effects of subgrouping occur in OSS. To this end, we provide an initial, non-exhaustive assessment of faultline types in OSS development.

By communicating through electronic means, members of OSS projects have limited possibilities to observe characteristics of their peers [8]. Faultlines based on *demographic attributes* may not be perceivable and thus irrelevant–unless members include demographic information in their public profiles. In fact, demographic attributes have not been found to be prominent among members [23].

Motivations to join OSS projects are manifold and thus harbor potential for splitting groups along identity-based faultlines. It could, however, be the case that like-minded individuals cluster in homogeneous groups. Motivations have been found to differ also based on project characteristics, i.e. size [24], which then would attract a specific type of developer. Since membership may not be fully determined by a single motivating factor and projects may cater to more than one need, e.g. enabling learning and at the same time providing opportunities to build reputation, motivation is likely to lead to identity diversity and thus faultlines in OSS projects.

Experience in OSS development in general and the specific project is expected to present an information-based faultline. Differences in professional experience are documented as faultlines [25], additionally in the context of OSS distinct differences in knowledge, which arguably is related to experience, are described [8].

Reputation as an individual's social status is important in the social fabric in OSS [26]. Reputation as the congruence of promised actions and actual behavior [27] is multi-faceted such that positive views in technical aspects can be coupled with negative social evaluations. Given this multidimensionality and basis for authority

[8], reputation is likely to differ for an individual between projects and for individuals within one project. These differences are expected to lead to a hierarchical structure and thus resource-based subgroup [11].

Differences in *activity type* are well-established differentiators in OSS projects and thus a likely faultline item. First, users and developers differ, where users mostly consume and at most make small contributions such as bug reports or minor changes, whereas developers contribute all major code advancements [8]. Within the set of developers, a hierarchy consisting of a core and more peripheral developers has been described: A set of core developers has a disproportionate share of contributions, which entails more influence and reputation, whereas a large number of peripheral developers contributes relatively little code [15].

The *extent and persistence of contributions* is another potential faultline. The overall amount of activity is expected to be an influential member characteristic. Abstracting from the specific contribution behavior, the core-periphery structure of OSS projects [15] is based on the extent of contributions. Activity is an antecedent to previously discussed characteristics such as experience and hierarchy. In addition, by contributing continually, members can build knowledge, which is a key criterion for advancing to more central roles [15]. Drawing on research into other open collaborative processes, roles are expected to be identifiable but flexible over time [28]. A subsequent reduction in activity may thus demote members from the core to peripheral contributors.

3 Hypotheses

We propose a set of hypotheses to investigate the correlation of the relative size of subgroups based on high reputation, issue focus, high contribution extent, and high contribution persistence and success of OSS projects as defined by community size.

3.1 Success in OSS

Success in OSS is not dependent on a single characteristic. The multi-faceted nature of OSS success is evident from the proposition of frameworks to assess success based on diverse indicators [29]. Following previous application [29], we use the size of the non-core OSS project community as an indicator of its external success since contributions of peripheral members are valuable to maintain the project [8]. The onboarding mechanism, i.e. to integrate new developers into the project has been described to be a difficult issue in OSS [6]: Community size is thus apt to indicate how well a project cannot only garner attention but recruit contributors, who potentially can advance to the core team. Based on the preceding discussion of characteristics prone to lead to faultlines and subgroups, we derive hypotheses on the relations of a selected subset of bases for subgrouping that are deemed relevant for community size.

3.2 Hypotheses on the Configuration of Subgroups in OSS Core

For the hypotheses on the relations of the relative size of subgroups defined by faultlines, we draw on findings concerning the stable yet dynamically changing roles in open collaboration [28]. Discussions on the type of potential subgroups and their related effects draw on the typology suggested in [11].

Reputation has been described as an individual's social status in OSS projects [26]. Drawing on extant research in subgrouping, differences in reputation can be related to hierarchical differentiation, which can lead to negative outcomes as resource-based subgroups [11]. At the same time, without perceptions of unfair distribution, a hierarchy can facilitate information processing and thus aid group performance [30]–which relates to potentially positive information-based subgroups. In the context of OSS, reputation has been found to increase trust and satisfaction in members [27, 31]. Considering virtual teams, trust in turn has been observed to increase participation and community activity [32]. Reputation may also facilitate the progression from observing user to contributor with decision power [8] through satisfaction, which leads to participation intentions [27]. In addition, since we focus on the size of the peripheral community as dependent variable, the presence of high-reputation individuals in the project core may signal credibility [27], which may help to attract new members.

H1: The relative size of the subgroup of high-reputation individuals in an OSS project will be positively correlated with community size.

Issue Focus: Different activity backgrounds lead to the potential of information-based subgroups, which can be positive [11]. A large share of issue-focused contributors, i.e. with most activity in creating and commenting on issues, is expected to foster community size. Reporting issues is a known pathway to transition from user to core contributor [33] since issue reports require less specific technical and project-related knowledge than code contributions. With commenting also being part of issue focus, a large share of issue-focused members implies many members may still be starting out as contributors or many people are helping others into the project by sharing knowledge through comments. Core members commenting on issues of newcomers is comparable to mentoring, which has been found to aid onboarding [34]. In addition, a large subgroup based on such behavior may send positive signals of a collaborative culture to outsiders and consequently make the project more attractive.

H2: The relative size of the subgroup of issue-focused individuals in an OSS project will be positively correlated with community size.

Contribution Extent: While rather general, the extent of contributions in projects is expected to foster success and to generate follow-up activity. Similar to reputation, past contributions in a project act as an outside signal of activity and maintenance–as opposed to a majority of OSS projects that are effectively abandoned [3]. Such signals may sway outsiders to become part of the peripheral network. Activity in and of itself is positive in OSS, which is witnessed by an emphasis on practical work in core beliefs [5]. The importance of activity for community building is mirrored in the finding that updates on activity are a key reason for following other members [35]. In addition, for acquiring new casual contributors, a large share of highly active contributors can make it easier for newcomers to identify who to turn to and ask questions and whose work to study to overcome issues related to a lack of replies found in onboarding [6]. In this sense, a large share of members in a high-activity subgroup may foster efficient processing as a knowledge-based subgroup [11].

H3: The relative size of the subgroup of extensively contributing individuals in an OSS project will be positively correlated with community size.

Contribution Persistence: Analogously to contribution extent, we expect a large share of persistent developers to aid community size. The presence of persistent contributors shows a project is actively developed and thus increases its attractiveness. Past activity may inform future activity [36] and thus benefit future contributions. In addition, persistent developers may be easier to identify and the likelihood of responses are increased, which can address the onboarding issue of receiving no reply from core members [6].

H4: The relative size of the subgroup of persistently contributing individuals in an OSS project will be positively correlated with community size.

Control Variables: To control for systematic differences in OSS projects, we include project age and the existence of previous releases as control variables. Project age is used to control for lifecycle aspects [37] and to capture related effects such as integration in the OSS community, access to resources and progress [38, 39]. As a binary control, the previous existence of releases is used to control for projects that while being actively developed do not declare official releases.

4 Method

4.1 Sampling

Data on OSS projects was obtained from GitHub Archive and a copy of the GHTorrent data on Google BigQuery. We included projects that had at least 100 pull requests or at least 500 commits and at least 2000 comments between January 1st 2014 and August 31st 2017, yielding 6037 projects of which we drew a 10% random sample. Controlling for name changes, the sample contained 580 projects. Since success is expected to be the result of collaboration we followed extant OSS research [37] and applied a lagged structure: Independent variables are collected from a six month timeframe in the middle of the project lifetime, community size in the following six months and control variables in the preceding six months. Sample size was reduced to 482 based on a sufficient level of activity in the reference period and the removal of two outliers showing an extreme level of activity not representative of the majority of projects.

4.2 Operationalization of Measures

Community size as a measure of external interest and thus success, see section 3.1, is operationalized as the extended development community [29] and implemented as the count measure of individuals associated with the project without being part of the project core. *Reputation of developers* is calculated as the prestige actor proximity index [40] by measuring how connected and how close an individual is to other members. This is a more elaborate approach than the one used by [26] to assess OSS reputation. Links between members are inferred by analyzing the sequence of users' comments, their quotes and direct references in discussions. The index increases with the number of reachable developers and if developers, which are directly or indirectly connected, get closer. *Contribution extent* is operationalized as an individual's share of overall project activity during the period of investigation, in terms of comments, issues, commits, and pull requests. This measure is inspired by previous constructions of developer-activity pairs in networks [2]. *Persistence* of an individual is operationalized as the share of periods with activity in all periods since the individual's first contribution. Following previous research classifying contribution types in OSS [2], *issue focus* shows the relative share of issue-related activity in an individual's contributions. It is operationalized as the ratio of the number of issue-related contributions to a project, e.g. issue reporting and commenting, compared to an individual's overall contributions to the project. We were, however, unable to reliably distinguish comments from issues to those to pull request, which may skew results.

Table 1 Operationalization of Measures

Measure	*Operationalization*	*Calculation*
Community size [Success]	Extended Development Community of p as sum of non-core members i_p^{nc}	$S_p = \sum i_p^{nc}$
Reputation	Connectedness and closeness of developer i to other developers j through comments, quotes, and direct references	$P_{ip} = \frac{I_i/(g_P - 1)}{\Sigma_j d(j,i)/I_i}$ I_i Number of nodes reachable from i g_P nodes in project p $d(j,i)$ distance of j to i
Contribution Extent	Share of i of overall activity in project p	$CE_{ip} = \frac{C_{ip}}{\Sigma_i C_{ip}}$ C_{ip} *Number of Contributions of i to p*
Contribution Persistence	Share of periods with activity since initial activity.	$CP_{ip} = \frac{A_{ip}}{S_{ip}}$ A_{ip} *# of periods with activity of I in p* S_{ip} *# of periods since first contribution of i to p*
Issue Focus	Share of issue-related activities in an individual's contributions	$R_{ip}^{issue\ focus} = \frac{C_{ip}^{issue}}{C_{ip}^{total}}$ C_{ip}^{issue} *# of issue-related contributions of i to p*

4.3 Analysis

For the independent faultline measures, we restricted analysis to the core project based on an activity threshold of twenty contributions. Since the count-based raw scores are prone to project-specific skews, values of faultline measures are normalized first. Drawing on previous findings concerning the specific effects of the relative size of subgroups [10], we operationalize the size of theoretically derived subgroups as the share of team members deviating more than half a standard deviation from the project median. Core members of projects can thus belong to either the high or low-value subgroup on the respective measure. This calculation is done for each independent variable. With the dependent variable being a count measure, Negative Binomial Regression (NBR) has been chosen as regression method.

5 Results

Table 2 Correlations of Variables

	High Reput.	Issue Focus	Contrib. Extent	Contrib. Persistence	Proj. Age	Comm. Size
High Reputation	1					
Issue Focus	-0.21***	1				
Contrib. Extent	-0.29***	0.29***	1			
Contrib. Persistence	-0.13***	0.16***	0.20***	1		
Project Age	-0.03	0.07	0.10**	0.26***	1	
Community Size	0.09*	0.00	-0.17***	0.10**	0.08*	1
Note:	*p<0.1; **p<0.05; ***p<0.01					

Correlations of variables–shown in Table 2–are relatively small with the maximum value being .29 in absolute terms. The direction of correlations is, however, worth mentioning: the share of members with high reputation is negatively correlated with all other measures, whereas all other correlations are positive. Table 3 details the regression models. The first model only includes the subgroup measures as independent variables, whereas the second and third one add the control variables project age and whether there have been releases in the project. The control variables do not change the direction of correlations but influence levels of significance. While we expected the attribution of members to subgroups to be meaningful for community size, the information content of the model is rather low judging from the pseudo R^2 values.

H1 regarding the effect of a large share of members with high reputation is partly supported: We observe a positive albeit insignificant and small relation.

H2 regarding the effects of a large share of members with a focus on issue activity is likewise partly supported with a small positive, albeit insignificant relation.

H3 regarding the effect of a large share of extensively contributing members is not supported with a highly significant and strong negative relation.

H4 regarding the effect of a large share of persistently contributing members is supported with a significant and strong positive relation.

Table 3 Results of Regression Models

	Community Size (Number of Non-Core Members)		
Model	**1**	**2**	**3**
Relative Size of Subgroup of Members with			
High Reputation	.621*	.411	.413
Issue Focus	.377	.307	.317
Contribution Extent	**-2.176*** **	**-1.926*** **	**-1.959*** **
Contrib. Persistence	**1.283*** **	**1.304*** **	**1.201*** **
Releases		**.491*** **	**.484*** **
Project Age			.001
Constant	**4.797*** **	**4.482*** **	**4.171*** **
Pearson Dispersion	1.257	1.228	1.26
Pseudo R^2 (McFadden)	0.007	0.011	0.011
Pseudo R^2 (Nagelkerke)	0.076	0.117	0.121
Observations	482	482	482
Log Likelihood	-2,762.340	-2,751.393	-2,750.319
theta	.804*** (.046)	.834*** (.048)	.837*** (.048)
Akaike Inf. Criterion	5,534.680	5,514.786	5,514.639
Note:	*p<0.1;	**p<0.05;	***p<0.01

6 Discussion

Based on the characterization of work in OSS projects and theory on faultlines and subgrouping, we investigated the relation of contribution-based subgroups and community size as a measure of success. While the applied regressions explain only a small share of variance, large sets of core members with high reputation, a focus on issues, and especially persistent contributions positively relate to community size. A large share of extensively contributing members is significantly negatively related.

As expected, a large share of high-reputation core members has a positive but small and after including controls insignificant relation with community size. With reputation being a key aspect of OSS culture [5], we expected the resulting differences in power, resource access and status to be attenuated by the culture of OSS work and thus to lead to a positive relation. Results suggest, however, that reputation may also in OSS projects lead to negative repercussions–possibly due to an identity- or resource-based subgroup. This relation may interact with the hypothesized positive effect. Operationalized as a social proximity measure, the positive finding is in line with previous work on the positive effect of internal cohesion for OSS success [37]. Structurally speaking, a larger share of developers with more direct access to other core members positively relate to community size. Drawing on previous research describing a positive effect of loosely coupled, decentralized developers for design as opposed to technical work [2], the expected signaling effect of high-reputation projects may draw in peripheral contributors–whereas the technical work of closely related core contributors may trigger feelings of inaptitude and thus present barriers to onboarding [6, 41]. Previous findings thus help interpret the only partly expected findings. In addition, reputation based on the distance of the social network may not be perceivable to outsiders and thus reduce the expected signaling effect. The current operationalization of reputation may overestimate values of members being in constant exchange with others without adding value to the project. Investigating other metrics, e.g. formal collaborator status, thus seems worthwhile.

The positive relation between a larger share of members focusing on issue activity and community size is relatively small and insignificant. With a grain of salt, this result may be interpreted as slight proof of the proposition that issue-focused core members foster community size as defacto mentors helping to overcome onboarding issues [34]. Moreover, supporting others as a core value in OSS [5] could propel membership. This line of reasoning has, however, to be questioned since the correlation between the share of high-reputation and issue-focused individuals is slightly negative. The small effect size may be due to our specific threshold values for considering members part of the project core as it could have included too many contributors and thus left no room for outside community. In addition, as stated before, the operationalization of issue focus suffers from the inability to classify some comments. Effects may be more reliably tested if the content of contributions was to be analyzed in more detail: In particular, the community building effect may be identifiable if responses to activity by non-core members were studied in particular.

The relations of contribution extent and persistence are somewhat surprising: Persistence is–as expected–significantly positively related, while extent is significantly negatively related. The correlation between the two measures is weak, which implies they capture distinct contribution types. Persistence might not only be the sum of contributions over time but may signal activity, future maintenance, and thus value in contributions to outsiders. It might also imply technical proficiency and learning forming part of the OSS culture [5]. Seeing sense in one's contributions can be related to the intrinsic motivational factors as key drivers of OSS membership [1, 4]. Persistence might thus signal a project is worthy of contributions. Extensive contributions on the other hand are operationalized by the overall activity of individuals. A large share of extensively contributing core members might create the impression of a closed circle and thus deter contributions–relating to the finding that newcomers face barriers in where to start contributing [6]. In addition, the negative correlation of contribution measures and high reputation is noteworthy since it may imply that contribution quality by high-reputation individuals is distinct from quantity.

This research provides initial evidence that the relative size of subgroups in OSS projects may have differential effects based on underlying faultlines. Findings on positive relations add to existing research on the onboarding of new members [6, 34] by outlining potential levers for action. It seems plausible that persistence and issue-related work can act as mentoring and thus as means to help newcomers get started [34]. For OSS practitioners, analyzing and possibly steering the observed relations may be helpful for increasing community size and thus potential human capital in their projects. Results also add to the discussion on the effects of balanced versus imbalanced configurations of subgroups [10, 11]. Our findings indicate that a larger share of members as an imbalanced configuration may have positive outcomes. Results further add to research on the differential effects of subgroups depending on their reason for formation and typology [11, 21].

7 Limitations and Future Research

This research is only a first step towards understanding the configurational properties of OSS members based on faultline and subgroup theory. There are several limitations, which in part may also explain the low pseudo R^2 values. A significant set of limitations arises from the choice of sampling and model specification. First of, the filtering criteria for including projects may have skewed results. In addition, the choice and operationalization of variables affect results. As faultline and subgroup measures, the entire breadth of characteristics studied in group research and psychology are conceivable. While carefully developed, operationalizations may not capture the phenomenon under study as expected. As an example, the operationalization of success as community size is just one option considering propositions to operationalize OSS success in multiple dimensions [29]. Furthermore, the inability to classify some comments may have skewed results. We strongly encourage further research using additional variables and testing the applicability of other operationalizations. Data has been collected during a limited timeframe using a lagged structure, which may have reduced explanatory power, especially if the timeframe studied was not representative for longer running projects. To inch closer to causal inferences, other methodologies such as experiments or mixed method approaches may be beneficial.

The distinction between core members and community poses two issues. Firstly, the threshold of attribution of members to either group may skew results. The operationalization may conflict with the observed fluid nature of OSS teams [23]. Secondly, we have studied relations of the share of members in the core team on the size and the extended community. This implies that all effects are indirect across the boundary of the core team, which may have caused some unexpected results.

Addressing these limitations and investigating additional aspects are promising avenues for future research. The effects of faultlines and subgrouping in virtual, loosely coupled groups warrants further exploration. In addition, investigating the effects of subgroups on outside individuals seems promising. Moreover, it may be worthwhile to investigate interaction effects between the proposed subgroups. As an example, the share of members with a combination of high reputation and issue focus would further the investigation of the proposed onboarding mechanism provided by these factors.

8 Conclusion

Open Source development has become an established organizational way of building software. Performance effects of faultlines and subgroups are commonly discussed in team research. While most subgrouping is described to be detrimental for team outcomes, more recent works have proposed to consider different basis for subgrouping and their configuration. We have investigated the faultline-based subgroup concept in OSS projects by first identifying characteristics that may trigger faultlines and subgroups. This assertion is the basis for the empirical investigation of the relations of community size and the relative share of core members belonging to subgroups characterized by high levels of reputation, focus on issues, and extent as well as persistence of activity. We find significant relations with the size of the extended project community by contribution persistence and extent and positive albeit insignificant relations of a large share of members with high reputation and issue focus. Our results add to extant research on subgrouping and their configurational properties. In addition, they provide an additional step towards understanding how success as community size of OSS projects can be fostered.

9 Acknowledgements

This research and development project was funded by the German Federal Ministry of Education and Research (BMBF) within the Program Innovations for Tomorrows Production, Services, and Work (02K14A080) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

References

1. von Hippel, E., von Krogh, G.: Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. Organ. Sci. 14, 209–223 (2003).
2. Temizkan, O., Kumar, R.L.: Exploitation and exploration networks in open source software development: An artifact-level analysis. J. Manag. Inf. Syst. 32, 116–150 (2015).
3. Khondhu, J., Capiluppi, A., Stol, K.-J.: Is It All Lost? A Study of Inactive Open Source Projects. In: OSS 2013, IFIP AICT 404. pp. 61–79 (2013).
4. Hertel, G., Niedner, S., Herrmann, S.: Motivation of software developers in open source projects: An Internet-based survey of contributors to the Linux kernel. Res. Policy. 32, 1159–1177 (2003).

5. Stewart, K.J., Gosain, S.: The Impact Of Ideology on Effectiveness In Open Source Software Development Teams. Mis Q. 30, 291–314 (2006).
6. Steinmacher, I.. b, Graciotto Silva, M.A.., Gerosa, M.A.., Redmiles, D.F..: A systematic literature review on the barriers faced by newcomers to open source software projects. Inf. Softw. Technol. 59, 67–85 (2015).
7. Lau, D.C., Murnighan, J.K.: Demographic diversity and faultlines: The compositional dynamics of organizational groups. Acad. Manag. Rev. 23, 325–340 (1998).
8. Daniel, S., Agarwal, R., Stewart, K.J.: The Effects of Diversity in Global , Distributed Collectives: A Study of Open Source Project Success. Inf. Syst. Res. 24, 312–333 (2013).
9. Thatcher, S.M.B., Patel, P.C.: Group Faultlines. J. Manage. 38, 969–1009 (2012).
10. O'Leary, M.B., Mortensen, M.: Go (Con)figure: Subgroups, Imbalance, and Isolates in Geographically Dispersed Teams. Organ. Sci. 21, 115–131 (2010).
11. Carton, A.M., Cummings, J.N.: A theory of subgroups in work teams. Acad. Manag. Rev. 37, 441–470 (2012).
12. Howison, J., Crowston, K.: Collaboration Through Open Superposition: a Theory of the Open Source Way. MIS Q. 38, 29-A9 (2014).
13. Wiesche, M., Krcmar, H.: The relationship of personality models and development tasks in software engineering. In: Proceedings of the 52nd ACM conference on Computers and people research - SIGMIS-CPR '14. pp. 149–161 (2014).
14. Sahraoui, S., Al-Nahas, N., Suleiman, R.: An Agile Perspective on Open Source Software Engineering. Int. J. Open Source Softw. Process. 4, 52–65 (2012).
15. Crowston, K., Shamshurin, I.: Core-periphery communication and the success of free/libre open source software projects. J. Internet Serv. Appl. 8, (2017).
16. Tan, Y., Mookerjee, V., Singh, P.: Social capital, structural holes and team composition: Collaborative networks of the open source software community. In: ICIS 2007 Proceedings (2007).
17. Grewal, R., Lilien, G.L., Mallapragada, G.: Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems. Manage. Sci. 52, 1043–1056 (2006).
18. Jehn, K.A., Bezrukova, K.: The faultline activation process and the effects of activated faultlines on coalition formation, conflict, and group outcomes. Organ. Behav. Hum. Decis. Process. 112, 24–42 (2010).
19. Przybilla, L., Wiesche, M., Krcmar, H.: The Influence of Agile Practices on Performance in Software Engineering Teams: A Subgroup Perspective. In: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research. pp. 33–40. ACM, New York, NY, USA (2018).
20. Faraj, S., Sproull, L.: Coordinating Expertise in Software Development Teams. Manage. Sci. 46, 1554–1568 (2000).
21. Carton, A.M., Cummings, J.N.: The impact of subgroup type and subgroup configurational properties on work team performance. J. Appl. Psychol. 98, 732–758 (2013).
22. Pflügler, C., Wiesche, M., Krcmar, H.: Subgroups in Agile and Traditional IT Project Teams. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018).
23. Vasilescu, B., Filkov, V., Serebrenik, A.: Perceptions of diversity on GitHub: a user survey. In: Proceedings of the Eighth International Workshop on Cooperative and Human Aspects of Software Engineering. pp. 50–56 (2015).
24. David, P.A., Shapiro, J.S.: Community-based production of open-source software: What do we know about the developers who participate? Inf. Econ. Policy. 20, 364–398 (2008).

25. Bezrukova, K., Jehn, K.A., Zanutto, E.L., Thatcher, S.M.B.: Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance. Organ. Sci. 20, 35–50 (2009).
26. Bird, C., Gourley, A., Devanbu, P., Swaminathan, A., Hsu, G.: Open borders? Immigration in open source projects. In: 29th International Conference on Software Engineering Workshops(ICSEW'07) (2007).
27. Casaló, L. V., Cisneros, J., Flavián, C., Guinalíu, M.: Determinants of success in open source software networks. Ind. Manag. Data Syst. 109, 532–549 (2009).
28. Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., Gurevych, I.: Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. Inf. Syst. Res. 27, 792–812 (2016).
29. Crowston, K., Annabi, H., Howison, J., Masango, C.: Towards a portfolio of FLOSS project success measures. In: Workshop on Open Source Software Engineering, 26th International Conference on Software Engineering (2004).
30. Meyer, B., Glenz, A., Antino, M., Rico, R.: Faultlines and Subgroups : A Meta-Review and Measurement Guide. Small Gr. Res. 45, 633 –670 (2014).
31. Antikainen, M., Aaltonen, T., Väisänen, J.: The role of trust in OSS communities - Case Linux Kernel community. In: IFIP International Federation for Information Processing. pp. 223–228 (2007).
32. Ridings, C.M., Gefen, D., Arinze, B.: Some antecedents and effects of trust in virtual communities. J. Strateg. Inf. Syst. 11, 271–295 (2002).
33. Herraiz, I., Robles, G., Amor, J.J., Romera, T., González Barahona, J.M.: The processes of joining in global distributed software projects. In: Proceedings of the 2006 international workshop on Global software development for the practitioner - GSD '06 (2006).
34. Fagerholm, F., Münch, J., Borenstein, J., Park, M.: The Role of Mentoring and Project Characteristics for Onboarding in Open Source Software Projects. In: ESEM'14 (2014).
35. Blincoe, K., Sheoran, J., Goggins, S., Petakovic, E., Damian, D.: Understanding the popular users: Following, affiliation influence and leadership on GitHub. Inf. Softw. Technol. 70, 30–39 (2016).
36. Ouellette, J.A., Wood, W.: Habit and intention in everyday life. Psychol. Bull. 124, 54–74 (1998).
37. Singh, P., Tan, Y., Mookerjee, V.: Network Effects: The Influence of Structural Capital on Open Source Project Success. MIS Q. 35, 813–829 (2011).
38. Subramaniam, C., Sen, R., Nelson, M.L.: Determinants of open source software project success: A longitudinal study. Decis. Support Syst. 46, 576–585 (2009).
39. Stewart, K.J., Ammeter, A.P., Maruping, L.M.: Impacts of license choice and organizational sponsorship on user interest and development activity in open source software projects. Inf. Syst. Res. 17, 126–144 (2006).
40. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994).
41. Bandura, A.: Social foundations of thought and action. Englewood Cliffs, NJ. (1986).

The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance

Tim Rietz[1], Ivo Benke[1] and Alexander Maedche[1]

[1] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, Karlsruhe, Germany
{tim.rietz,ivo.benke,alexander.maedche}@kit.edu

Abstract. Information technology is rapidly changing the way how people collaborate in enterprises. Chatbots integrated into enterprise collaboration systems can strengthen collaboration culture and help reduce work overload. In light of a growing usage of chatbots in enterprise collaboration systems, we examine the influence of anthropomorphic and functional chatbot design features on user acceptance. We conducted a survey with professionals familiar with interacting with chatbots in a work environment. The results show a significant effect of anthropomorphic design features on perceived usefulness, with a strength four times the size of the effect of functional chatbot features. We suggest that researchers and practitioners alike dedicate priorities to anthropomorphic design features with the same magnitude as common for functional design features in chatbot design and research.

Keywords: Acceptance, Anthropomorphism, Chatbot, Collaboration, Work

1 Introduction

Recently, the usage of chatbots for improving collaboration in the workplace has seen increasing interest [10]. Chatbots create new opportunities in digital work, potentially boosting collaboration culture [25]. Additionally, they have the potential of positively influencing the balance between the time allocated to work and private activities through digital interventions [5] and reducing work overload through supporting task management [45]. Major collaboration platforms in and outside the workplace include chatbots, such as Facebook, Slack, WhatsApp and Telegram. Slack has established itself as a successful platform used by thousands of companies, due to its capabilities for group collaboration and native integration of various productivity tools. Collaboration is commonly defined as making a joint effort toward a group goal, where joint effort encompasses acts of shared creation and/or discovery [3]. Collaboration systems used in the working context are referred to as enterprise collaboration systems, with Slack as a prime example. Airbnb, Autodesk, IBM and many others [40] frequently use the collaboration features of Slack and the possibility to access over 1000 chatbots developed by professionals and freelancers alike [34], alongside Slack's

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

natively integrated chatbot 'Slackbot'. As chatbots in enterprise collaboration systems are expected to become a substantial element of the modern workplace, there is a need to better understand the impact of chatbot design on user acceptance.

We observe a lack of user acceptance studies discussing overall chatbot design features in collaborative environments. In the face of limited resources, developers are required to make trade-off decisions regarding two essential aspects when designing chatbots: form vs. function. Traditionally, engineering-oriented disciplines tend to pay more attention to the functional dimension and dedicate less resources to the form dimension [7]. Form describes the relationship between design parameters and is primarily perceived as an aesthetic expression [36]. Due to its hedonic nature, form has a strong link to social presence of a bot and has been shown to positively affect perceived enjoyment and ease-of-use [17, 37]. This paper investigates the influence of anthropomorphic and functional chatbot design features in enterprise collaboration systems on user acceptance, on the basis of importance and frequency of usage of design features in Slack by practitioners. We formulate the following research question:

"How do anthropomorphic and functional chatbot design features in enterprise collaboration systems influence user acceptance of chatbots?"

In order to answer the research question, we follow a survey-based research approach and specifically investigate user acceptance of chatbots in the context of the enterprise collaboration system Slack. We contribute to the chatbot design body of knowledge by investigating how different design features influence user acceptance. At the same time, we provide a contribution for practitioners involved with chatbots in collaboration environments by providing input for the form vs. function trade-off decision in chatbot design.

2 Theoretical Background

2.1 Chatbots in Collaboration Systems

Chatbots are applicable for a large variety of situations, such as supporting collaborative learning [43], but are simpler in development and interaction compared to complex intelligent agents. Design decisions have to be made with regards to the 'chatting behavior' of the agent. Various configurations of chatbots are used in research and practice, such as agents that react dynamically to changing environments [51] or aid only on invocation (e.g. Slackbots /remind me functionality). Furthermore, seemingly small details of their conversational behavior, such as social cues, have a significant influence on the agents' effect on users [50]. Social cues are applied as an anthropomorphic feature of an agent, resulting in increased social presence [1]. In the example of MentorChat, a web-based agent for collaborative learning support, collaboration between individuals was enhanced by triggering discussions between students [43]. Furthermore, studies have shown that collaboration in a professional context suffers from a lack of group leaders, and that conversational agents can act as digital replacement for these roles [9].

2.2 Form and Function of Chatbots

Design science literature describes design features as concrete ways of integrating design principles into artifacts [28]. In the context of this paper, we use the term *design feature* as name for a group of functionalities that chatbots may provide. We refer to these individual functions as *items*. Chatbot design draws from two dimensions: form and function, the two fundamental components of design across domains [44], which definitions are displayed in table 1. Despite arguments for considering additional concepts in design [23], form and function are extensively considered as complete in forming the design dimensions [9].

Table 1. Literature definitions of the form and function design dimension

Design Dim.	*Definitions in Literature*	*Source*
Form	[…] while form refers to certain customer interface characteristics and is often addressed from the perspective of visual aesthetics.	[2]
	We define form as structural product characteristics that provide the architecture through which functional product features are delivered. Product form embodies the hedonic component of design.	[44]
	The form of the object as a whole can then be represented as the collection of components and a description of the interaction among components.	[36]
	[…] alternatively, design has been equated to product form, focusing on its esthetic characteristics. This approach has generally found that these attributes are related to hedonic value.	[23]
Function	Function refers to certain product function characteristics and their perceived performance […]	[2]
	Product function refers to product specifications and standard architectures - essentially the utilitarian aspect of product design. Functional design is defined by the factors, benefits, characteristics, and features that are combined to provide utility.	[44]
	[…] functional requirements which describe performance. There are many designs which satisfy any one set of functional requirements, therefore there cannot be a unique relationship between the function and the form of a product.	[36]

Form. The form of a product or service on the other hand refers to the arrangement of individual design components [36]. It is primarily perceived as an aesthetic expression [9] and can be interpreted as a user’s perception of non-utilitarian aspects. Form features mostly are hedonic in nature and characterized through pleasure derived from the appearance of a product / service [33]. Historically, form features of design are investigated in marketing and product development literature [25], and comprise a multitude of elements, such as usage of lines, curves, proportions and symmetry. In the domain of websites and software, research regularly focuses on visual aesthetics when discussing form features [24]. In the design of chatbots however, another form feature becomes a relevant research topic: anthropomorphic presentation of the virtual agent [38], (cf. chapter 4.2). Anthropomorphism is considered part of the form design dimension, as its items change the visual presentation of an agent and the interaction between components. Although the most commonly used variant of anthropomorphic virtual agents are embodied conversational agents [8], chatbots as well can incorporate anthropomorphic features. Despite a chatbot being limited in the range of applicable visual cues to appear more human-like, it may still rely on language that is enriched by emotional semantics or expression of emotions through emojis [42].

Function. The function of a product or service refers to parameters related to its general performance [36]. Historically, the function is dominated by principles from engineering [9]. The focus lies on providing utilitarian value, through addressing the practical needs of users. These can appear in simple functionalities, such as being able to communicate with an agent in natural language, or in more complex desires, e.g. safety or maintainability. Therefore, improving functional design features of a product/service pays consideration to how objects can be arranged in a way for users to interact with them efficiently and comfortably [44].

2.3 User Acceptance of Chatbots

We rely on a variation of the technology acceptance model (TAM) to investigate the impact of chatbot design features. Instead of utilizing the original TAM from Davis (1989), or its extensions, e.g. TAM2 [49], we decided for a model that includes perceived enjoyment (PE) as a core antecedent of behavioral intention [17]. The focus on chatbots in a work context stresses the interplay of hedonic and utilitarian aspects of a system [17]. Utilizing chatbots for improving company-internal collaboration might still be perceived as a novel application by employees [32]. This may introduce an hedonic component into work processes, potentially contributing to user acceptance of information systems (IS) normally perceived as utilitarian [17]. Besides PE, we do not include further concepts into our TAM evaluation, such as social presence or trust, as it is not in the scope of our evaluation. The study is intended to use a lean research model with a limited number of concepts. We are aware of concerns with the TAM methodology regarding its generalizability [4], the inclusion of factors external from the technology, or the models capabilities for predicting IT adoption [26]. Consequently, we evaluated the applicability of TAM as basis of the proposed research model against other common models used in IS research, such as UTAUT [47] or Task-Technology Fit [13]. The simplicity of the model, a broad range of research on its primary constructs alongside the inclusion of PE provides us with an appropriate model for this initial assessment. The traditional measures PEoU, focusing on the degree of peoples believe that using an IS is free of effort, and PU, focusing on the expected increase in job performance through using an IS [48], allow us to explore the utilitarian antecedents of behavioral intention. PE, the degree to which fun can be derived from using the IS, on the other hand explores the hedonic antecedents of it [17].

3 Research Model

Anthropomorphic form features influence users' perceptions of social presence, which in turn has been shown to positively affect PE [1, 33]. We expect the same relationship to be demonstrated in the context of chatbots in Slack. Furthermore, previous studies on TAM showed the close linkage between PEoU and hedonic values [17, 37]. As the form feature of design primarily induces hedonic value, we expect a positive relation between anthropomorphic chatbot design features and PE and PEoU.

H1: Anthropomorphic chatbot design features positively influence the perceived enjoyment of chatbots in enterprise collaboration systems.

H2: Anthropomorphic chatbot design features positively influence the perceived ease-of-use of chatbots in enterprise collaboration systems.

As we observe work context, the conceptual distinction between hedonic and utilitarian concepts may blur, since hedonic systems would more naturally occur in home environments [17]. As such, we expect to observe effects of the anthropomorphic design feature on PU.

H3: Anthropomorphic chatbot design features positively influence perceived usefulness of chatbots in enterprise collaboration systems.

A product design study identified PEoU as the most cited benefit (40%) of the utilitarian design [23]. Another study, investigating the role of function in utilitarian design of mobile data services, found a significant effect of function on user satisfaction, positively affecting PEoU and PU [2]. Thus, we formulate the following hypothesis for the effect of function features on TAM constructs:

H4: Functional chatbot design features positively influence perceived ease-of-use of chatbots in enterprise collaboration systems.

H5: Functional chatbot design features positively influence perceived usefulness of chatbots in enterprise collaboration systems.

Finally, in accordance with H1, hedonic and utilitarian concepts may blur at work and so we expect to observe effects of functional design features on PE.

H6: Functional chatbot design features positively influence perceived enjoyment of chatbots in enterprise collaboration systems.

Additionally, we investigate the effects of selected control variables – age, gender, experience (with Slack, with Slackbots, general developing experience, and chatbot developing experience in Slack) and education level of the participants. Figure 1 depicts the resulting research model.

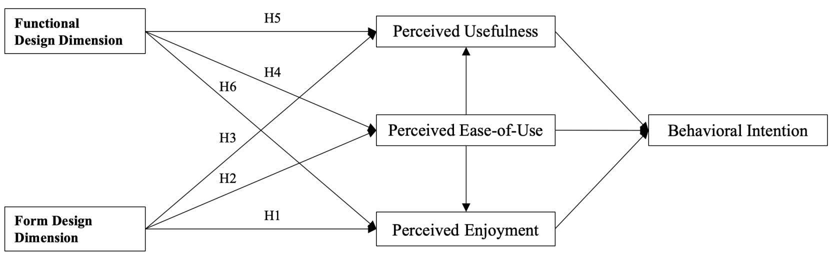


Figure 1. Research Model

4 Empirical study

4.1 Context: Enterprise Collaboration System Slack

As the experiment investigates design features of chatbots, the form of communication is text-based and the common technology for interacting with chatbots is instant messaging. Because we explicitly want to explore the effect of different design features on technology acceptance in a work context, we chose Slack as platform for our study. Slack is widely used in international enterprises and therefore provides a suitable solution for conducting a study. Slack has integrated a default chatbot, called "Slackbot". This default bot is approachable within every conversation. It can be enhanced by individually implemented commands. This function makes Slack particular interesting for this experiment. Besides its application in the work context, the natural integration of self-deployed chatbots by providing an interface API is necessary for providing custom designed solutions.

4.2 Selected Chatbot Design Features of Function and Form in Slack

Relevant design features are selected through a multi-stage process. Initially, a list of possible interactions, functionalities and behaviors available to bots in Slack is extracted from the Slack API documentation [41]. From the complete set of bot capabilities in Slack, we select five distinct groups of capabilities, whose effect on behavioral beliefs we attempt to measure. The five groups refer to design features of chatbots in Slack, that are commonly observable in bot implementations. The features are categorized according to the definitions of design features and its dimensions provided in table 1. Design features where pre-tested for importance and completeness with eight doctoral candidates familiar with using Slack in a work context and interacting with chatbots. During the pre-test the participants were confronted with the feature instantiations and the questionnaire. Their feedback was collected, evaluated and merged in order to develop the final survey. The following features are used as instantiations of the **functional dimension**:

Invocation. Usually, there are two ways of engaging in a conversation with a chatbot: either it is invoked by a command or it reacts by activating autonomously. In Slack, the user may activate the 'reminder' functionality of a chatbot by invoking the /remind command. At the same time, when pasting links into a conversation, e.g. from Google drive, the Slackbot will recognize the link and autonomously suggest a specific reaction to every subsequent link.

Intention-type. When interacting with a chatbot, the interaction can revolve around two intentions, seeking information or delegating a task [45]. A user may ask the bot for an update on the weather or to write an apology message to another user. More advanced chatbots may also provide a combination of both tasks, such as Google Assistant making a reservation with a service provider while sending information back to the user [27].

Question-type. The more sophisticated the interaction with a chatbot becomes, the more natural and simpler may the conversation feel. However, a sophisticated interaction increases the bot's complexity [18]. In order to keep complexity low, a chatbot will ask closed questions (such as Yes/No questions). This form of interaction allows the chatbot to dictate the flow of the conversation. Open questions on the other hand allow the user to communicate in a natural, unrestricted way.

Reaction-type. Finally, chatbots have various strategies of handling failure. When a bot struggles to understand or act on a command, it may react with an error notification and ask the user to retry or provide additional information. Other fallback mechanisms may include forwarding a user query to a common search machine or asking for more information. This behavior could be characterized as an attempt of the chatbot to satisfy the user in a fallback scenario by relying on outside sources and providing links [39].

The following feature is used to instantiate the **form dimension**:

Anthropomorphism. In presenting the bot to its users, a common approach is the humanization of the conversational agent, especially present in embodied conversational agents. For chatbots, the possibilities for making the bot appear as a human-like actor are limited. Common approaches are the inclusion of social cues in conversations, such as including short pauses before giving an answer [23], or making the bot include emojis in their conversational behavior [42]. The underlying idea of making agents more human-like stems from the CASA paradigm – computers are social actors [31, 35]. The paradigm states that social attributes are ascribed to computer technology during interaction, similar to another human, and was proven with sets of studies testing e.g. mindless responses in detail and the depth of social responses to computer's personality [31]. This can result in the development of social and emotional bonds with the agent [22] Social presence was identified as an antecedent of behavioral intention through affecting the PE of users [20]. Furthermore, the principle serves as a possible explanation for the observation that group allow agents to take on social roles, such as group leader, and viewing a machine as an embodied entity by referring to the agent as "him" and not "it" [8]. In Slack, chatbots can have a user profile, providing the same information as the profile of a human user. For explanatory reasons this aspect of anthropomorphic design features is displayed in figure 2.

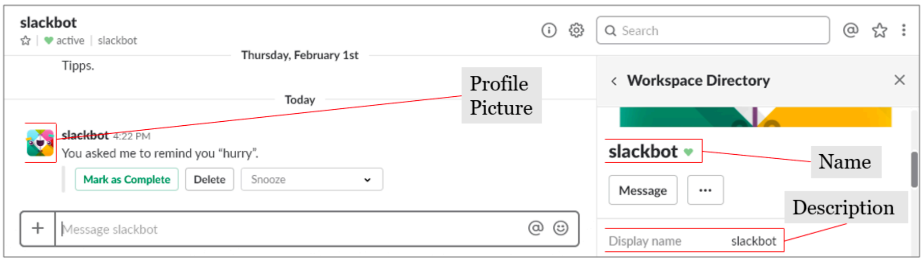


Figure 2: Picture and Name of Chatbot as Example of Anthropomorphic Design Feature

4.3 Procedure & Participants

This study is conducted as a confirmatory survey study. The study is set up using Limesurvey, a web application for creating and distributing research surveys. The survey language is English. At the end of the survey, participants have the opportunity to participate in a lottery for the chance of receiving rewards. For the purpose of eliminating survey replies not completed truthfully, we included two reverse coded and two trap questions. Trap questions ask participants to select specific answers to ensure reading the questions properly. For the evaluation procedure, the finalized survey is sent out to enterprises using Slack by posting the survey with an introduction text into relevant Slack workspaces. For this purpose, we have access to Slack workspaces of three international companies in the area of consulting, eCommerce & fashion and IT industry. Additionally, Slack has invitation only workspaces for professional chatbots, where the survey is distributed as well. We record a total of 71 answers. From those answers, 14 are not fully completed out or failed to provide the appropriate answer to included trap questions and therefore are excluded. In consequence the study sample contains 57 participants (N=57), whereby 48 are male (84,3%) and 9 are female (15,7%).

4.4 Measurement Approach

The structural model consists of two parts, the hedonic TAM and function and form dimensions. The constructs of TAM are derived from [17]. All items of TAM are measured on a seven-point Likert scale ranging from "Highly disagree" to "Highly agree". The question and answer forms are individually composed according to [46]. The design dimensions of chatbots in Slack are measured utilizing the design features introduced in section 4.2. Each feature is represented by two or more items, representing specific instantiations of the feature. To assess the influence of each item, the questionnaire applies a two-question characteristic, measuring importance and frequency of usage for each item. Both Importance and Frequency of usage are measured on a seven-point Likert scale. The measurement scales are derived from [46]. The two question characteristic is adopted from [30], who applied the design to study the relationship between features of a game and we-intentions. We adopt a reflective path model, suggesting that both importance and frequency of usage are a sample of the possible indicators for the respective latent construct. For the calculation of the two latent constructs, functional and form design dimension, we additively combined the two characteristics, importance and frequency of usage, for each item. We controlled for age, gender, education and experience. Experience is measured on four levels: Slack experience, chatbot experience, chatbot experience in Slack and general coding and development experience, on a seven-point Likert scale ranging from "Not at all experienced" to "Extremely experienced", as derived from [46]. An exemplary set of questions is depicted in table 2.

Table 2. Exemplary question set for one item from the Invocation design feature

Measurement characteristic	*Question*
Importance	How important is it to you, that the chatbot can be tagged in conversations like human users?
Frequency of usage	How often do you invoke the chatbot via a user command to perform some action?

5 Results

To evaluate the collected survey data, we apply partial least squares (PLS), a structural equation modelling (SEM) technique [7]. Specifically, we use the tool SmartPLS 3.2.7. As we try to explore relationships and the proposed research model, traditional common factor based PLS-SEM is chosen over component-based SEM [12, 14]. This approach is particular suitable for our model, as it is applicable for smaller sample sizes [14]. Moreover, we apply the bootstrapping method to calculate the significances of path coefficients.

5.1 Measurement Model

In a first step, we evaluate the measurement model. We apply a reflective instead of a formative modeling approach as the measurement items are representatives of the latent variable rather than a composition [14]. To assure the quality of our data, we determine the items' loadings and assure indicator reliability and validity (see table 3). All items of the measurement model load above 0.6 on their construct, which is an sign for indicator reliability [6]. Composite reliability (CR) is chosen to overcome limitations of Cronbach's α for measuring internal consistency. It reveals values at least between 0.7 and 0.8 for five out of six, a reliable indicator for an adequate confirmatory model and values higher than 0.8 for four of six constructs, indicating a good fit for confirmatory research for these constructs [14]. However, Cronbach's α for PE indicates almost no inter-item correlation, while CR for PE indicates correlation, but only at levels adequate for exploratory research [14]. The convergent validity was measured by average variance extracted (AVE). Except for one (AVE = 0.471), AVE shows values for all constructs higher than 0.5, which is sufficient [14]. However, as this represents only a small discrepancy from the accepted value for this one value, it is a minor limitation of the model but the analysis can be continued and is still valid [15]. Further, we assess discriminant validity by running the Fornell-Larcker criterion, confirming that the square root of the AVE exceeds the respective constructs correlation with other variables in the model [11].

Table 3. Reliability Measures of the Measurement Model

Latent Variable	*α*	*CR*	*AVE*	*Fornell-Larcker Criterion*					
				FuDD	*BI*	*FoDD*	*PEoU*	*PE*	*PU*
Functional Design Dimension (FuDD)	0.675	0.823	0.608	**0.780**					
Behavioral Intention (BI)	0.944	0.964	0.899	0.578	**0.948**				
Form Design Dimension (FoDD)	0.786	0.842	0.437	0.444	0.341	**0.661**			
Perceived Ease of Use (PEoU)	0.677	0.797	0.505	0.471	0.630	0.432	**0.711**		
Perceived Enjoyment (PE)	-0.094	0.617	0.639	0.488	0.535	0.255	0.662	**0.799**	
Perceived Usefulness (PU)	0.950	0.960	0.802	0.720	0.598	0.410	0.630	0.669	**0.895**

5.2 Structural Model

After evaluating the measurement model validity, we assess the structural model according to [17]. Figure 2 shows the structural model with the results of the PLS bootstrapping analysis with 5000 samples displaying the coefficient of determination (R2) for all endogenous constructs. The figure contains path coefficients, significance levels and effect sizes (f^2) of the paths.

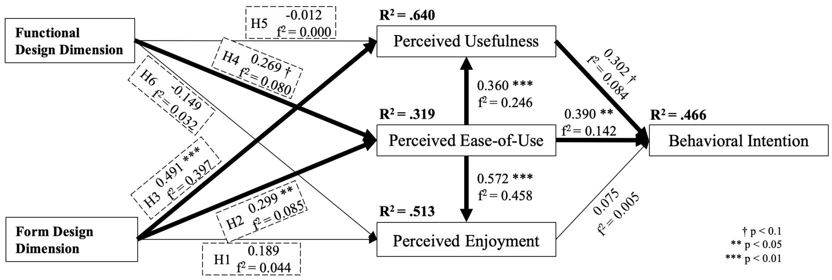


Figure 3. Structural Equation Model

We report significance on the following significance levels: $p < 0.01$, $p < 0.05$, $p < 0.1$. We explicitly allow for significances at the 10% level, as this study aims to identify promising correlations for future research to explore with more detailed and strict criteria, rather than minimizing false positives. For H1 our evaluation shows a significant effect of functional → PEoU ($p < 0.1$, $f^2 = 0.08$). H2 and H6 are not supported by the results as there is no significance measurable, whereas H4 and H5 are supported by the evaluation with significance at level $p < 0.05$ and $p < 0.01$, respectively. H3 is rejected as well, showing no significance. The constructs PU, PEoU and PE have positive loadings on BI. The paths from PEoU → PU and PEoU → PE also show a significant effect. Besides the path PE → BI, all of them are significant at least at the $p < 0.05$ level while only one is at the level of $p < 0.1$. The three outgoing paths from the anthropomorphic design features have a positive effect while the path to PU has a very strong path coefficient. The path to PEoU loads weaker. The effect size of anthropomorphic → PEoU (H4) counts for $f^2 = 0.085$, the value for anthropomorphic → PU (H5) for $f^2 = 0.397$. It can be concluded, that most of the paths have effect sizes $> f^2 = 0.02$, which at least accounts for small effects [16]. Only for PE → BI, the effect size is neglectable ($f^2 = 0.02$). The coefficient of determination for all constructs was medium up to high [7], with values between 0.319 and 0.64. Additionally, we test for the moderating effects of experience with developing, chatbots and Slack, but do not identify any significant effects.

6 Discussion

In contradiction to our initial hypothesis, anthropomorphic design features show no significant effect on PE. This rejects H3 and contradicts the assumption that form features have direct effect on the hedonic character of the acceptance of chatbots. Originating from the point of view that higher anthropomorphism is leading to higher perceived enjoyment of the user, this result is especially valuable. Drawing implications from this, it shows that a stronger humanization of chatbots does not necessarily result in higher user enjoyment. Additionally, no effects of significance involving PE, other than PEoU on PE, were observed in the study. A significant relationship between PE and BI, observed by previous studies (29), did not occur. Alongside the insignificant effects of form and function design features on PE, this may hint at the questionnaire failing to capture PE of participants. Cronbach's α for PE indicates no inter-item correlation, while CR appears at the lower end of the acceptable range. Consequently, the underlying construct items might not have measured the construct appropriately. On the other hand, the results may also be interpreted as PE not being a relevant factor to influence users' behavioral intentions, stressing the importance of the utilitarian character of chatbots in work environments. Taking a deeper look, all the more interesting are the further results of the anthropomorphic design features. They show the most significant effect of outgoing paths to PU. In addition, the path loading is strongest, and effect size is highest ($f^2 = 0.397$) of all evaluated paths. This means that the effect of anthropomorphic design features on PU is the strongest across all results. The finding is supporting H5. This result comes together with a significant path loading from anthropomorphic design features to PEoU, which supports H4. Although the effect is not as strong as to PU, it still is significant with a positive path loading and effect size considered as meaningful ($f^2 = 0.085$). Comparing the two paths - anthropomorphic design features → PU and → PEoU, we see the former one has almost twice the loading and effect size, showing a strong prominence of effect on PU.

Accompanying these results and contrasting them to the first discussion point mentioned above spawns a valuable implication. Instead of influencing the hedonic share of technology acceptance, anthropomorphic design features have the strongest and most significant effect on utilitarian aspects of chatbot acceptance. This was contrary to the initial assumption about humanization of chatbots and may therefore be seen as the main contribution of this research work. Further elaborating this idea, it means that Slack users accept a anthropomorphic chatbot more for utilitarian reasons than for joy and hedonic perception. As possible explanation, we call on the argument of context influence raised by van der Heijden [29]. It imposes that work environments are mainly associated with utilitarian values. The influence of anthropomorphic design features, which are assumed to be hedonic, seem to be controversial in this context. Our results suggest that they impact the utilitarian character of acceptance which goes hand in hand with the implication of [29], suggesting that hedonic features add acceptance to utilitarian systems and thereby usefulness, especially in the working context. Furthermore, we may theorize that user satisfaction is increased when interacting with chatbots that utilize anthropomorphic design features, thus benefiting PU and PEoU, similar to the effects that Botzenhardt et al. observed for functional design features [2].

Looking at the functional features of design characteristics, the only significant connection is pointing at PEoU. This supports H1. Therefore, we can report a positive effect of functional features on the ease-of-use of the chatbots technology and see the support of utilitarian aspects by functions in the design of a chatbot. On the other hand, H2 was not supported. This is interesting as well, suggesting that functional features may make usage easier but not necessarily make the chatbot more useful. Overall, we still document much higher loading, significance and effect size on the outgoing paths from anthropomorphic features than functional ones. Moreover, it is worth mentioning that H6 is rejected as well, stating no confirmative effect on PE. Mirroring the hypothesis about functional features (H1, H2, H6), our results at least suggest, that they are more likely to contribute to utilitarian aspects than to hedonic aspects, as we cannot confirm any significant relationships with PE.

Finally, aggregating our findings, they stress the importance of anthropomorphism in order to increase acceptance. This is valuable for both researchers and practitioners, giving a clear guideline for chatbot design and research. Observing this implication, it poses the questions if more anthropomorphism directly leads to higher acceptance. For answering this question, we want to mention the phenomenon of "uncanny valley" [29]. It compares human likeness of an entity with human affinity to this entity and states that the relationship is increasing, but only until reaching a specific point, the uncanny valley. Simultaneously, the raised equation is finally likely to fail, as humans are feeling rather unfamiliar when a technical system reaches this specific point of anthropomorphism.

6.1 Implications

"The aesthetics of your product speaks out for you just as much as the functionality, because if the functionality is no longer unique, guess what steps in? The beauty." [21]

We urge researchers and practitioners alike to pay attention to the importance of anthropomorphism for chatbot adoption. The results of this study show strong and significant effects on the utilitarian aspects of acceptance of the users. Moreover, taking up on findings on chatbots, they possibly serve as leaders in groups, the results gain even more significance as replacing a human being might demand for serving different social and hedonic facets. The importance of functional features goes without saying and the relevance of function features for PEoU could be demonstrated for the Slack in the work context. However, function and form features need to go hand in hand to achieve the best possible outcome. While function and form serve separate causes in the design of an agent, both are relevant for behavioral intentions of users. Tailoring both features to the individual task of a chatbot can significantly improve its acceptance amongst users.

6.2 Limitations

Some limitations apply to this study. The selected design characteristics for function may be perceived not as functionalities provided by a chatbot, but rather as characteristics of the agent's interaction with a user. As such, they may provide both utilitarian and non-utilitarian value. Furthermore, a specific chatbot task may affect the perception of function and form features. However, the selected function characteristics are present in chatbots fulfilling a large spectrum of tasks and purposes. Choosing distinct functions for this study might limit the findings of this study to specific fields of application. Regardless of these considerations, we suggest the evaluation of our results with chatbots dedicated to discrete tasks and purposes.

The study is framed as especially focused on the workplace. We assume this focus reveals different aspects than it does in private life. Though, there might be concerns that the tool Slack as well induce private mindset into the survey's results. We did not especially check for this in the questionnaire as we think it would not have fully covered the limitation. Regardless of these considerations, we suggest the evaluation of our results with chatbots dedicated to discrete tasks and purposes. Furthermore, we did not evaluate potential effects of the participants' PEoU, PU and PE on the design features proposed. As such, we cannot exclude reverse effects. Finally, despite applying PLS-SEM as evaluation method due to the small sample size, and adhering to the rule of thumb of 10 participants per [14], repeating the evaluation with a larger sample size may be advised to further solidify the results.

7 Conclusion

With this study, we evaluate the impact of anthropomorphic and functional chatbot design features on user acceptance in Slack, an enterprise collaboration system. We conduct an online survey, asking users of the common collaboration and messaging tool Slack about function and form features of chatbots. Contrary to our hypotheses formulated against the backdrop of previous studies on conversational agents, we identify anthropomorphism to have a highly significant effect on PU. The effect size of anthropomorphism on PU was four times the size of other significant effects identified. On these bases, we reject our initial hypothesis which predicted a strongly positive effect of form features on PE. These findings highlight the importance of form features, in the form of anthropomorphism, in achieving a high PU and consequently strengthening the BI of users. The results have implications for developers of chatbots in collaborative work environments. We urge researchers and practitioners alike to dedicate resources to form features in the same magnitude as dedicated to function features during chatbot development. At the same time, we encourage further research on the effect of function and form features of design in the context of specific chatbot tasks or specific collaboration environments, as well as other collaboration setups.

References

1. Al-Natour, S., Benbasat, I.: The Adoption and Use of IT Artifacts: A New Interaction-Centric Model for the Study of User-Artifact Relationships. J. Assoc. Inf. Syst. 10, 9, 661–685 (2009).
2. Botzenhardt, A. et al.: The Roles of Form and Function in Utilitarian Mobile Data Service Design. J. Electron. Commer. Res. 17, 3, 220–238 (2016).
3. Boughzala, I. et al.: Special Issue Team Collaboration in Virtual Worlds: Editorial to the Special Issue Team Collaboration in Virtual Worlds: Editorial to the Special Issue. J. Assoc. Inf. Syst. 13, 10, 714–734 (2012).
4. Bouwman, H., van de Wijngaert, L.: Coppers context, and conjoints: a reassessment of TAM. J. Inf. Technol. 24, 2, 186–201 (2009).
5. Cameron, G. et al.: Towards a chatbot for digital counselling. Br. HCI Conf. 2017. 1–7 (2017).

6. Chin, W.W. et al.: Advancing the theory of adaptive structuration: The development of a scale to measure faithfulness of appropriation. Inf. Syst. Res. 8, 4, (1997).
7. Chin, W.W.: The Partial Least Squares Approach to Structural Equation Modeling. Presented at the (1998).
8. Derrick, D.C. et al.: Automated group facilitation for gathering wide audience end-user requirements. In: Proceedings of the Annual Hawaii International Conference on System Sciences. pp. 195–204 (2013).
9. Domermuth, D.: The Pedagogy of Form versus Function for Industrial Design. (2009).
10. Frommert, C. et al.: Using Chatbots to Assist Communication in Collaborative Networks Christian. In: IFIP Advances in Information and Communication Technology. (2018).
11. Fuller, M.A. et al.: Efficacy in Technology-Mediated Distributed Teams. J. Manag. Inf. Syst. 23, 3, 209–235 (2007).
12. Gefen et al.: Editor's Comments: An Update and Extension to SEM Guidelines for Administrative and Social Science Research. MIS Q. 35, 2, iii (2011).
13. Goodhue, D.L., Thompson, R.L.: Task-Technology Fit and Individual-Performance. MIS Q. 19, 2, 213–236 (1995).
14. Hair, J.F. et al.: A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publ. 2, (2016).
15. Hair, J.F. et al.: Multivariate Data Analysis. Pearson Education Inc., New Jersey (2009).
16. Hair Jr, J.F.: A primer on partial least squares structural equation modeling (PLS-SEM), https://www.amazon.com/Partial-Squares-Structural-Equation-Modeling-ebook/dp/B01DT3F33I/ref=sr_1_1?s=books&ie=UTF8&qid=1495311972&sr=1-1&keywords=9781483377469, (2013).
17. van der Heijden, H.: User Acceptance of Hedonic Information Systems. MIS Q. 28, 4, 695 (2004).
18. Hill, J. et al.: Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. Comput. Human Behav. 49, 245–250 (2015).
19. Hong, J.H. et al.: ConaMSN: A context-aware messenger using dynamic Bayesian networks with wearable sensors. Expert Syst. Appl. 37, 6, 4680–4686 (2010).
20. Jin, S.-A.A.: The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment. Comput. Human Behav. 26, 3, 443–451 (2010).
21. Kim, M.: Beautiful and functional interfaces – UX Planet.
22. Komiak, Benbasat: The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. MIS Q. 30, 4, 941 (2006).
23. Kumar, M., Noble, C.H.: Beyond form and function: Why do consumers value product design? J. Bus. Res. 69, 2, 613–620 (2016).
24. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. Int. J. Hum. Comput. Stud. 60, 3, 269–298 (2004).
25. Lebeuf, C. et al.: How Software Developers Mitigate Collaboration Friction with Chatbots. (2017).
26. Legris, P. et al.: Why do people use information technology? A critical review of the technology acceptance model. Inf. Manag. 40, 3, 191–204 (2003).
27. Leviathan, Y., Matias, Y.: Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone, https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.
28. Meth, H. et al.: Designing a requirement mining system. J. Assoc. Inf. Syst. 16, 9, 799–837 (2015).

29. Mori, M. et al.: The uncanny valley. IEEE Robot. Autom. Mag. 19, 2, 98–100 (2012).
30. Morschheuser, B. et al.: Gamified crowdsourcing: Conceptualization, literature review, and future agenda. Int. J. Hum. Comput. Stud. 106, May, 26–43 (2017).
31. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. J. Soc. Issues. 56, 1, 81–103 (2000).
32. Pickard, M.D. et al.: Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? Comput. Human Behav. 65, 23–30 (2016).
33. Qiu, L., Benbasat, I.: Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. J. Manag. Inf. Syst. 25, 4, 145–182 (2009).
34. Recast.ai: (Almost) Every Slack and Facebook Messenger Bot - Universe, https://airtable.com/universe/expJrvnYEz5BooYKk/almost-every-slack-and-facebook-messenger-bot?explore=true.
35. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places. (1996).
36. Rinderle, J., Hoover, S.P.: Function and form relationships : strategies for preliminary design. (1990).
37. Rodrigues, L.F. et al.: Does ease-of-use contributes to the perception of enjoyment? A case of gamification in e-banking. Comput. Human Behav. 61, 114–126 (2016).
38. Seeger, A.-M. et al.: When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents. Spec. Interes. Gr. Human-Computer Interact. 15, 1–6 (2017).
39. Segura, C. et al.: Chatbol , a chatbot for the Spanish “La Liga.” In: International Workshop on Spoken Dialog System Technology 2018 (IWSDS). pp. 1–12 (2018).
40. Slack: Customer Stories | Slack.
41. Slack: Enabling Interactions with Bots | Slack, https://api.slack.com/bot-users.
42. Stoeckli, E. et al.: Exploring Affordances of Slack Integrations and Their Actualization Within Enterprises – Towards an Understanding of How Chatbots Create Value. In: Hawaii International Conference on System Sciences 2018 (HICSS-51). pp. 2016–2025 (2018).
43. Tegos, S. et al.: MentorChat: Introducing a Configurable Conversational Agent as a Tool for Adaptive Online Collaboration Support. In: 2011 15th Panhellenic Conference on Informatics. pp. 13–17 IEEE (2011).
44. Townsend, J.D. et al.: Form and function: A matter of perspective. J. Prod. Innov. Manag. 28, 3, 374–377 (2011).
45. Toxtli, C. et al.: Understanding Chatbot-mediated Task Management. (2018).
46. Vagias, W.: Likert-type scale response anchors, (2006).
47. Venkatesh, V. et al.: User acceptance of information technology: Toward a unified view. MIS Q. 27, 3, 425–478 (2003).
48. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. 39, 2, 273–315 (2008).
49. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field. Manage. Sci. 46, 2, 186–204 (2000).
50. Wang, W. et al.: Effects of Emoticons on the Acceptance of Negative Feedback in Computer-Mediated Communication. J. Assoc. Inf. Syst. 15, 8, 454–483 (2014).
51. Wünderlich, N. V., Paluch, S.: A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents. Proc. Int. Conf. Inf. Syst. 1, 1–11 (2017).

Digital Feedback for Digital Work? Affordances and Constraints of a Feedback App at InsurCorp

Emanuel Stoeckli[1], Falk Uebernickel[1], Walter Brenner[1], Andrea Weierich[2], and Sarah Hess[2]

[1] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
{emanuel.stoeckli,falk.uebernickel,walter.brenner}@unisg.ch
[2] InsurCorp Technology, Munich, Germany
andrea@weierich.de, hess.sarah@outlook.com

Abstract. Little is known about how digital work shapes the exchange of performance feedback, even though today's digital and global world demands for more continuous feedback than annual reviews. This research investigates a feedback app in a naturalistic context within a globally leading financial service corporation (InsurCorp). Drawing on malleability and voluntary participation, the app offers possibilities to send and request feedback between employees. Rich contextual insights from a multinational pilot study with 568 users are gained by triangulating qualitative data from 21 semi-structured interviews and 69 feedback app user reviews with usage data. Anchored in the theory of affordances, we provide insights on use practices and find that the app affords operational-level feedback exchange on specific subjects, while general feedback on sensitive topics is preferably exchanged in person. To understand actualization facilitators and barriers, we take a social-technical systems perspective to elaborate contextual factors that influence the individual's actualization decision.

Keywords: Performance Feedback, Digital Work, Affordance Theory.

1 Introduction

Exchanging performance-related feedback on work is key to ensure individual and organizational progress [1, 2]. Scholars and practitioners agree that the ongoing digitalization is changing the nature of work [3–5]. Accordingly, the question arises how the rise of digital work is shaping the exchange of performance feedback [3, 4, 6].

First, traditional performance management processes such as once-a-year goal settings, performance reviews and 360-degree feedback are losing their appropriateness for the twenty-first century. They are typically long, lack in visible outcomes and are less valuable than conversations that take place in the moment of performance [7, 8]. Accordingly, besides assessing the performance from a retro-perspective, there is a shift towards individualized real-time feedback that guides future action and facilitates improvement, training and development [8, 9]. Second, in today's digital and globalized world, virtual, distributed and remote work settings demand for digital work tools [10, 11]. In fact, the majority of knowledge workers relies on digital technologies

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

[3, 4, 6], e.g., 83 percent of employees in Germany use digital technologies at work [12]. Accordingly, novel digital technologies not only offer opportunities for knowledge workers to perform work, but also to exchange feedback [8].

Prior research shows how motivation and productivity can be improved through altering the likelihood of receiving feedback [13], providing computer-generated feedback [14–16], providing real-time feedback on specific behavior [17], and embedding feedback features into task-specific collaboration environments [18]. However, there is a lack of research that investigates novel digital work tools dedicated to facilitating performance feedback exchange between employees. Accordingly, calls for research emphasize the need to examine the use of technology for performance management [8] and to investigate digital work tools that support knowledge workers in their digital work environments [3, 11].

Against this backdrop, we adopt a case study research strategy to investigate a digital feedback app and its use in the context of a pilot project in a naturalistic workplace setting at the global financial service provider InsurCorp. While the action possibilities offered by the feedback app may be perceived as enabling as well as inhibiting to employees (i.e., perception of affordances and constraints), employees continuously decide how to realize value from using the app (i.e., actualization). Therefore, we pose the following research question: *How is the perception and actualization of affordances and constraints from feedback apps affected by employees' individual use practices and organizational context factors?*

To do so, we apply a sociomaterial perspective to adopt "a relational middle ground between technological determinism and social constructivism" [19, p.2].

2 Theoretical Background

2.1 Performance Feedback in the Context of Digital Work

Sending and receiving feedback has become a key activity of knowledge workers to exchange information that relates to their performance and understanding [20, 21].

Traditional Performance Feedback. In this research, we focus on a particular form of feedback, that is, performance feedback. Drawing on prior research, we consider performance feedback as "dynamic communication process occurring between two individuals that convey information regarding the receiver's performance in the accomplishment of work-related tasks" [1, p.260]. Thereby, literature distinguishes formal and informal feedback. The former denotes official and top-down enforced events (e.g., yearly goal setting, performance appraisal and 360-degree reviews) [9], whereas the latter describes feedback events that take place independent of formal mechanisms during day-to-day work [22]. Opposed to formal events, informal events often have the advantage of being more timely and contingent on the situation of performance [1, 22]. This is important, because effective feedback is said to be timely (e.g., reducing feedback cycles), specific (e.g., related to a specific event/subject), relevant for the performer (e.g., enabling to request feedback), accurate, and easy to understand [23]. Accordingly, work usually involves both formal and informal feedback exchange through feedback seeking and giving [22].

Digital Work Context. Work is increasingly characterized as digital work, i.e., "[an] effort to create digital goods or that makes substantial use of digital tools" [6, p.283]. Consequently, the possibilities to assess performance and exchange feedback digitally are rising. On the one hand, the work of blue-collar as well as white-collar workers can be recorded and analyzed [14]. This enables the provisioning of computer-generated feedback, which is often preferred and more trusted by employees as it directs employees' attention to the task leading to higher performance [14–16]. Similarly, the availability of performance information enables the provision of real-time feedback while engaging in a particular behavior, thereby reducing salience bias and causing greater behavioral shifts than aggregated feedback [17]. In fact, changing the likelihood of receiving feedback improves productivity [13]. On the other hand, and aside from computer-generated feedback, feedback exchange between employees occurs digitally, i.e., computer-mediated feedback [15]. Rather than on platform-based digital work tools with embedded task-specific feedback mechanisms [11, 18], we focus on digital work tools that offer possibilities for employees to provide and seek feedback [7, 8, 24]. This is particularly relevant for understanding others' subjective judgments, e.g., for managers to assess their effectiveness [10]. A particular type of instantiation of such work tools are feedback apps, e.g., used by Amazon, Deloitte, GE, and IBM [7, 24].

Feedback App. We regard feedback apps as *digital work tools dedicated to providing employees with possibilities to exchange feedback in their day-to-day work* [8]. Similarly, social software creates interaction potentials for employees to exchange information [25, 26]. Conceptualizing feedback apps as social software emphasizes two key characteristics that inform our research. First, it is malleable and flexible, and hence, open to various yet unforeseen use contexts [27, 28]. In fact, malleability and flexibility are crucial for digital work design, because "human workers have individual, diverging, and continuously changing needs" according which digital solutions need to be adopted [3, p.2]. Second, it relies on voluntary participation and emphasizes bottom-up engagement instead of top down enforcement [29].

2.2 A Sociomaterial Perspective

Digital artifacts such as feedback apps entail forms of physical and digital materiality, which are relevant to users and endure across time and place [30]. However, to obtain meaning and effects from technological structures requires their enrollment in practices embedded in institutional contexts [31], e.g., in shared routines and hierarchies [28, 30]. Even though structure may endure across some time and place, neither technological nor social structures are fully stabilized and can change [28]. In fact, they are interdependent in that "(1) all materiality is social in that it was created through social processes and it is interpreted and used in social contexts and (2) that all social action is possible because of some materiality" [30, p.10]. A sociomaterial perspective acknowledges this interdependency by adopts a relational perspective.

Theory of Affordances as Theoretical Lens. Grounding this research in the affordance theory puts the emphasis on the perceived possibilities that objects offer to humans in a certain context [32]. Proposed in the domain of ecological psychology, the theory is widely adopted in IS research [19, 33, 34]. Its relational nature proves to be useful as it theorizes both, the human and the technical aspects of IS. This guides our research to mutually investigate the properties of the IT artifact (i.e., a feedback app), and the goals and capabilities of the users (i.e., employees within InsurCorp). Thereby, three conceptual distinctions shape the present research: *affordance emergence*, *perception* and *actualization* [19]. First, *affordance emergence* describes goal-oriented action potentials that arise from the relation between a specified user and a specific IT artifact [34]. Affordances are real so that – in our case - the possibility to request feedback with an app exists whether or not a user perceives or exploits it [32]. Second, *affordance perception* represents the recognition of action potentials. They may or may not be (mis-)perceived by a user. Both depends on factors such as available information [35]. Action possibilities offered by an IT artifact are not always enabling but may also be constraining depending on the user's goals and capabilities [28, 36]. This may trigger changes in technologies or in routines, which in turn, may lead to changed perceptions of affordances and constraints [28]. For instance, the same artifact that once was perceived as enabling by a user, may suddenly be perceived as constraining, because the goals have changed, or the use of the IT artifact showed that the goals cannot be achieved [28]. Third, *affordance actualization* describes the realization of actions potentials which, in turn, leads to effects. Technology simultaneously liberates and controls human action and is, thus, both constraining and affording to a certain extent; what dominates not only depends on the user, but also on the institutional context in which the user is situated and the technology is embedded [36]. Consequently, not only the emergence and perception, but also the actualization decision varies across contexts and depends on factors such as expected outcomes and perceived efforts required [35]. An artifact may provide employees with possibilities to fulfill their goals and they may perceive them, however, still they may decide to not actualize them. Employees compare affordances of artifacts with similar affordances, which other artifacts offer [37]. Hence, the actualization of affordances by individual employees not only depends on their goals and the artifact's materiality, but also on the organizational context in which an employee is situated and performs its daily work [30, 31, 33, 35–39].

Actualization as Socio-Technical Phenomena. Prior research relies on socio-technical systems (STS) theory to elaborate the sociotechnical conditions of affordance emergence and perception [40] as well as organizational changes required to actualize organizational-level affordances [41]. Both, the investigated feedback app artifact and employees are embedded in an organizational work context. Accordingly, we draw on STS theory [42, 43] and build on prior work [41, 44] to inform our investigation of the actualization process, because the alignment of the four STS components facilitates technology use, while gaps in alignment impede technology use [44]. STS theory understands organizations as systems of actors, structures, tasks, and technologies [42, 43]. Specifically, actors comprise people with qualifications [43]. Structures refer to systems of communication, authority, and work flow [43]. Tasks represent the "raison d'être: the production of goods and services, including the large numbers of different but operationally meaningful subtasks" [43, p.1144]. At last, technologies describe available means for problem-solving such as computers [43]. In summary, studying the actualization of affordances at individual level requires a broad recognition of the socio-technical context that may stimulate the actualization in varies ways [39].

3 Research Design

3.1 Case Setting

Striving to contribute towards theory development on performance feedback in the context of digital work, we inductively gain rich empirical data [45, 46] from a case study to investigate the phenomenon of interest in its real-world context [47]. Namely, a pilot project introducing a feedback app into a global financial services corporation.

Social Setting. The pilot comprises 568 participants situated in a naturalistic work environment at InsurCorp, which employs between 100 and 150 thousand employees and operates globally in the fields of insurance and asset management. There are three key stakeholders. First, InsurCorp's technology provider in Germany runs the pilot and has the vision to transform InsurCorp into a digital group. Second, the global human resources (HR) entity finances the pilot. Accordingly, the project team consists of a project manager, a product owner, and an intern of the first two stakeholders. Third, different operational entities introduce the app. The recruitment process started by consciously selecting entities based on location and specialization. InsurCorp Germany, France and Morocco were selected as national companies running the core business. InsurCorp Technology in Germany and Singapore were selected to include technology-oriented companies. Investment Management, Communication and Corporate Responsibility, and Global HR were selected due to their international orientation. Next, the HR responsible of each entity invited employees to participate on a voluntary basis considering both executives and non-executives as well as males and females from various job roles and departments.

Technological Setting. The introduced artifact is a customized app dedicated to the exchange of feedback (see Figure 1). It has two main features: sending and requesting written feedback or points. Each feedback must follow the structure "I like, I wish" and is non-anonymous. All personal feedbacks are listed in an inbox and the app shows a ranking based on the quantity of exchanged feedbacks. The app is not available in the company app store but is a separate mobile web app accessible via URL. However, it is accessible from everywhere. Due to works council agreements and the limited number of licenses, registration is compulsory and follows a manual workflow.

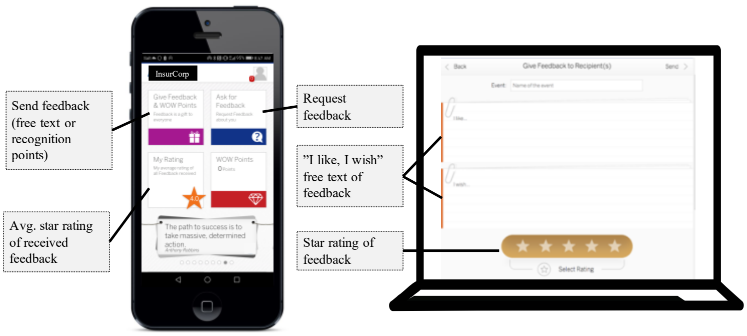


Figure 1. Screenshots of the feedback app on a mobile (left) and desktop (right) device.

3.2 Data Collection and Analysis

Data Collection. We obtained in-depth qualitative data from four sources of evidence and quantitative usage data from one additional source (Table 1) [47]. This suits well to address the sociomaterial and contextual research questions and given the restrictions of the works council. Similar to the recruitment of pilot participants, the interviewee recruitment process started by seeking names of pilot users from the HR responsible person of the individual entities. The goal was to cover executives and non-executives as well as heavy and light users (that used the app at least once) from varies functional and cultural backgrounds (see Table 1 and Table 2). To further disclose the interview process, all interviews were conducted either in German or English depending on the native tongue of the interviewee to increase the expressiveness of their statements. The interview guide follows well-established guidelines [48] and is grounded in the affordance theory. We started with questions to get to know the interviewees. Next, we focused on daily routines and work practices, since they "often oppose top-down specified production processes, and studying these processes creates a deeper understanding of individual needs" [3, p.4]. Accordingly, we continued with open-ended questions to prompt how and why the feedback app is used (or not) in everyday work. Further questions probed action possibilities and constraints as well as how these perceptions are influenced by the organizational context. Then, we shifted from today's use to changes over time and future use potentials. The interview guide was discussed within the project team and pre-tested in the first two interviews by analyzing the data and by requesting feedback using the app. The interviewees assessed the process with 4.5 out of 5 stars: "I liked the questionnaire as it allowed for some deeper evaluation of the use and potential of the tool. I think this is the correct way to collect feedback about the app at this stage. You also managed to create a pleasant trustful atmosphere which makes it easy to speak openly".

Table 1. Multiple sources of evidence

Data source	*List of details and descriptive statistics*
1. Qualitative semi-structured interviews	21 interviews with pilot participants (#1 to #21); 62% females / 38% males; 47.7% heavy users / 52.3% light users; 38% executives / 62% non-executives; Jul.-Dec. 2017; between 18 and 51 min. transcribed recordings.
2. Feedback app user reviews	69 reviews as answers to feedback requests using the feedback app (#FR); Aug. 2017 to Dec. 2017; 48% females / 52% males; avg. of 119 characters "I like" & 145 characters "I wish" statements; avg. of 3.82 of 5 stars.
3. Status meetings	Weekly WebEx calls within the project team; from Jul. 2017 to Mar. 2018; between 30 and 60 minutes.
4. Verification of results within InsurCorp	Discussion of results with the project team (face-to-face and WebEx), the manager responsible for people sourcing and development (face-to-face), the OE managers (WebEx), and with InsurCorp consulting (WebEx).
5. Quantitative usage data	Usage data (e.g., number of sessions, number of exchanged feedbacks, distribution of feedback ratings); Restricted by the works council to aggregated and transactional data (i.e., no feedback content).

Data Analysis. Following guidelines for qualitative research [46, 47, 49], data was iteratively collected and analyzed until a coherent picture emerged. We triangulated our sources of evidence in MAXQDA 12 [47] by adopting open, axial and selective coding [46]. The unit of analysis are individual employees within their organizational context, thus, analyzing individual affordances and constraints, while applying replication logic across operational entities [47]. During open coding, codes were assigned inductively to condense the transcripts. Our axial coding procedure was based on the theoretical lens to code (1) properties of the app, (2) properties of pilot users (i.e., goals, capabilities and context), and (3) (mis-)perceived and (non-)actualized affordances as well as constraints. During selective coding, we sharpened the connections between the affordances and constraints as well as the relations of the organizational context to the emerge-perception-actualization process. Drawing on related research [30, 40, 41], we extended the coding structure with the dimensions of socio-technical systems [42, 43] to elaborate factors of the organizational context in which an individual is situated.

Then, we triangulated our qualitative insights with quantitative data. First, over six months, 6,2% of users engaged in 26 to 50 sessions, while 45,9% of users only had one session. Second, feedback exchange decreased over time. Given the first month is 100%, the number of exchanged feedbacks decreased in the subsequent months: 65%, 23%, 22%, 5% and 4%. Third, the distribution of ratings associated with feedback is as follows: 5 stars (58%), 4 stars (33%), 3 stars (6%) 2 stars (2%), 1 star (1%).

Table 2. Characteristics and number of interview partners per operational entity and location

Entities	*Locations*	*Job roles (interview duration and type)*
Technology Provider (6)	Germany (2), Singapore (4)	#1 Head of Central Function Platforms (34 min, f2f) #11 Intern (37min, phone) #6 Human Resources Services (18min, phone) #7 Asia Core Systems (46min, phone) #8 Asia Core Systems (39min, phone) #14 Head of Tech. Prov. Singapore (27min, phone)
Investment Management Alpha (4)	Germany (3), Hong Kong (1)	#2 Employee Experience (51min, f2f) #3 HR Systems Consultant (47min, f2f) #4 Head of HR Digital (45min, phone) #5 HR Solutions Specialist (14 min, phone)
Investment Management Beta (3)	Germany (3)	#9 Head of Fixed Income (27min, phone) #19 Chief Investment Officer (37min, f2f) #21 Asset Liability Manager (49min, phone)
Global HR (2)	Germany (2)	#10 Processes - HR Transformation (23min, phone) #20 Head of People Sourcing & Dev. (38min, phone)
Insurance Morocco (3)	Morocco (3)	#12 Head of Dev. & Engagement (36 min, phone), #13 Portfolio Manager (28min, phone) #15 Audit Intern (RB, 25min, phone)
Communication (3)	Germany (3)	#16 Jun. Communication Manager (26min, phone) #17 Internal Communications Officer (35min, f2f) #18 Project Manager (40min, f2f)

4 Results

4.1 Affordances and Constraints of the Feedback App

Due to its malleability, the feedback app is open for wide variety of feedback. In practice, however, we identify common use practices in the form of four use scenarios and five use trajectories that explain how employees perceive and actualize the feedback app as a digital work tool to exchange feedback on concrete and operational activities, i.e. the first-order affordance.

Use Scenarios. First, the feedback app is used for *onetime activities*. For example, “I saw a presentation that I found particularly good, then I tried the app” (#7). Furthermore, the app was used to request feedback after meetings: “We did a lot of stuff around, we had a team offsite, I asked them [the participants] for their opinion on that.” (#2). Aside from group meetings, the app is used for one-on-one meetings: “I simply sent the people a request to give feedback or after giving a presentation to my boss” (#7). Second, use scenarios include *recurring activities*. For example, “we have a weekly call for [team name]. So, I asked ‘how do you like the weekly call?’ [and] ‘is it useful at all?’” (#3). Third, feedback is exchanged upon *phase changes* such as delivering projects, completing milestones, or finishing the first week at work. For instance, “when we delivered a project, then, of the four or five people working together, I would give some feedback to each of them” (#2). Fourth, the app is used to acknowledge *desired behavior*. “When something has been quite amazing, […], you say ‘that was good. that's quite nice’. You just want to give a little pat on the back.” (#17). Lastly, it was used to say thanks, e.g., “[to my manager for being] very calm and supportive and really helping me to be constructive” (#2).

Use Trajectories. Our results reveal how employees take up five trajectories of use. The first two use scenarios are rather typical (cf., Path 1-2 in Figure 2), while the three additional trajectories show how employees continuously navigate between physical and digital spaces to perform work and exchange feedback (cf., Path 3-5 in Figure 2).

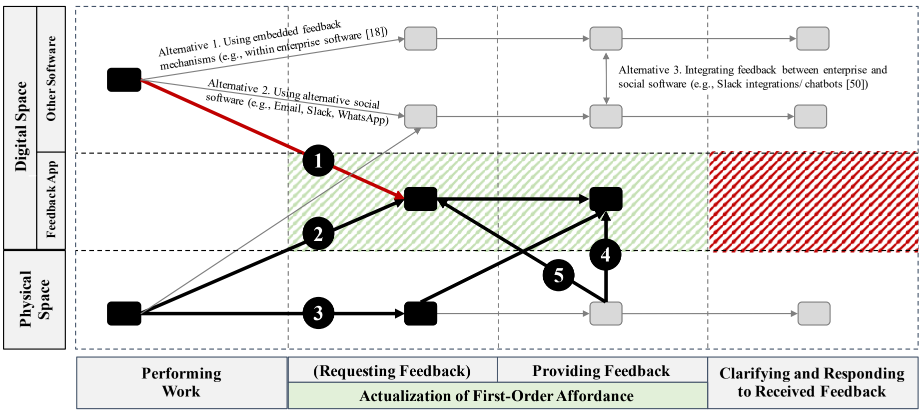


Figure 2. Trajectories (black) of actualization (green), alternatives (gray) and constraints (red)

First, employees exchange feedback on digitally performed work (e.g., a WebEx call). *Second*, the app is used for offline performed work (e.g., presentations) by either requesting feedback from colleagues or providing unsolicited feedback. *Third*, employees refer to the feedback app in conversations: “I don’t think I have done it [i.e., requesting feedback] by sending it from the app, but I rather asked them directly. More when we are in the dialogue… like you put in an additional sentence… is there any feedback feel free to use the feedback app” (#2). Accordingly, the app is not used until a colleague decides to actually send feedback. *Fourth*, even if feedback is exchanged in person, additional feedback may be exchanged afterwards. For instance, “yesterday we had a meeting with all the team leaders and there were some who had a comment or an idea after the meeting, […], you can do that via the feedback app.” (#1). *Fifth*, employees assess if they make progress on the feedback they received in person: “I had received feedback in different one-on-one discussions that I was trying to action, […], and so I asked some of them if I was moving the needle at all on that.” (#3).

First-order Constraints. *First*, constraints emerged from lacking integrations of the feedback app with other enterprise software (e.g., WebEx, Outlook). Employees perceived high media change efforts compared to alternative means to exchange feedback (see Figure 2), while the richness of the mediated feedback was limiting for some employees: “Can the app record snippets of a WebEx presentation?” (#FR). *Second*, employees perceived constraints from the limited possibilities to see who is registered and the impossibility to send feedback to non-registered co-worker. One employee explained: “I haven't used the app very much so far, since it is not very transparent who of my colleagues has signed up for the pilot phase” (#FR). It was argued that “if we could send feedback to someone that is not registered, it would push him/her to register” (#FR). *Third,* some employees expected to send anonymous feedback: “I thought this was anonymous and it doesn't seem like it is and that's a deal breaker.” (#17). *Fourth*, employees perceived constraints in clarifying and responding to feedback (see the red rectangle in Figure 2): “I would actually like to have more of an interaction on the feedback that I give… like a feedback on the feedback I’m giving… or […] that I’m receiving. […]. And when I receive a feedback where there is something positive and there is something constructive for improvement… then […] I want to answer to that” (#2).

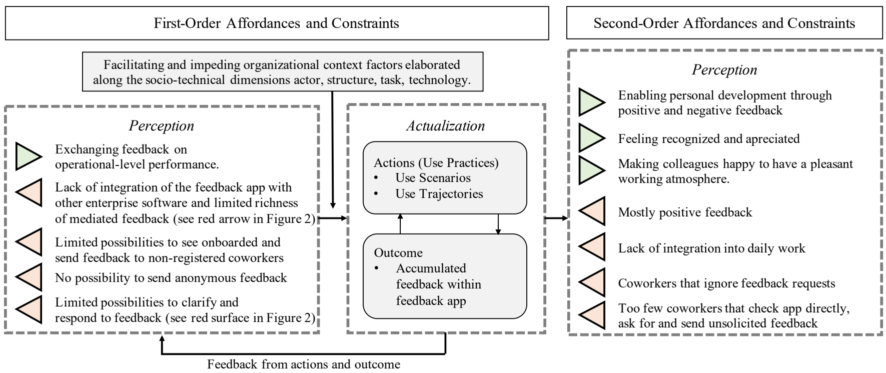


Figure 3. First-order and second-order affordances (green) and constraints (red)

Second-order Affordances and Constraints. Actualizing the first-order affordance to exchange operational-level performance feedback enables the emergence of second-order affordances and constraints (see Figure 3). Given the accumulated operational-level feedback as outcome of the first-order affordance, employees see potentials to for personal development by identifying weaknesses and strengths through positive and negative feedback: "it can actually give me more stuff to work on. To see what are my weaknesses and my strengths and how to improve overall" (#FR). As such, the app serves as "feedback account" to collect feedback in one place. In turn, constraints emerge from the way coworkers actualize the first-order affordance. Most notably, employees mention an emphasis on positive feedback: "Using the app, I realized most feedback remains personal and the app will be biased to only positive comments" (#FR). This is consistent with the rather high star rankings associated with the feedback. Interestingly, only some employees find this constraining, while others are satisfied and compare it to social media: "I do not write on anyone's Facebook wall 'I dislike your beach picture'. […] You can tell if your contributions are good in that if you get likes for it, it's probably good […, and otherwise] it was probably only average" (#21). However, employees consistently reported to feel happy and recognized when they received feedback: "It's a simple thing but receiving points or positive feedback really makes your day even better" (#FR). And sending feedback enables employees to maintain a pleasant working atmosphere that fosters motivation and a feeling of belonging together: "it just makes people feel good when you make them happy, and then it is more pleasant to work together" (#1).

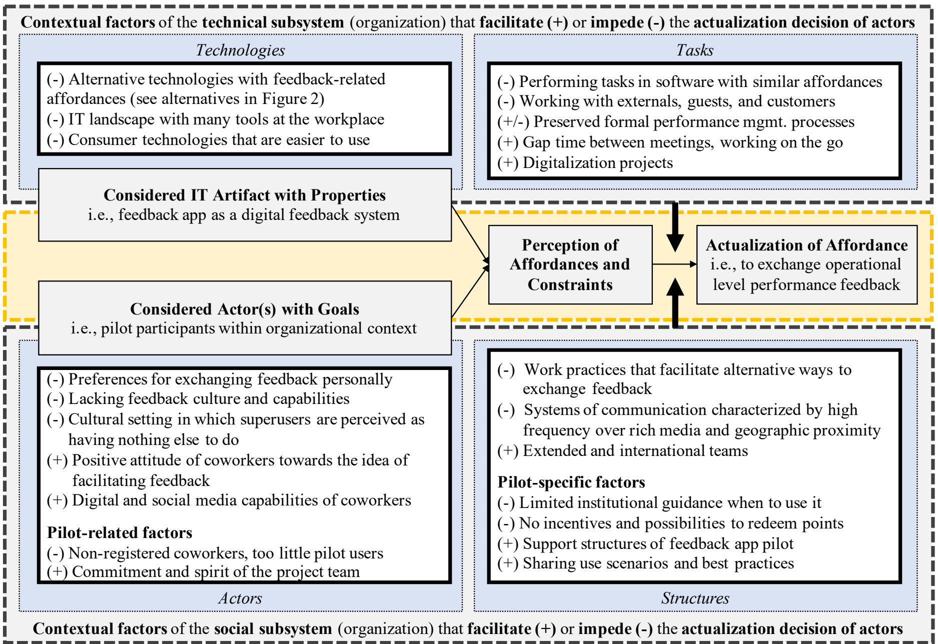


Figure 4. Facilitating and inhibiting context factors of the technical and social subsystem

4.2 Socio-Technical Context Factors that Facilitate and Impede Actualization

Still, employees may or may not realize the perceived possibilities to exchange feedback. In fact, the number of exchanged feedbacks decreased over time. Our results suggest that understanding individual's actualization decision requires considering their sociotechnical context, which comprises facilitating (+) and impeding (-) factors.

Technical Subsystem. The feedback app is part of a larger technical subsystem. Employees use it in the context of other technologies at work. A fragmented IT landscape with too many tools impedes actualization: "I wish NOT to work with an additional tool" (#FR). This goes along with alternative technologies that offer similar affordances, e.g., an employee stated that "for me it's actually equivalent to email. But not much better" (#9). Comparisons also include consumer technologies: "when you WhatsApp call people and it appears 'rate the quality of your call', you just click a star and then it [the feedback on the call] goes away" (#17). The technological context, in turn, goes along with the task environment. Performing tasks in software that offers feedback-related affordances inhibits employees' willingness to exchange operational-level performance feedback with the feedback app (see alternatives in Figure 2). Furthermore, working with externals, guests and customers is an inhibiting factor. For example, an auditor points out: "I used it only once, because I wasn't in the company. I am doing inspections, so I go around Morocco" (#15). Consequently, individuals expressed their need to "use it with guests/customers, because this feedback counts the most" (#7). Embedding the pilot in a task environment in which formal performance management processes are preserved was perceived as facilitating and inhibiting. On the one hand, it is additional work: "I would find it ideal if the feedback app is developed so that it replaces the 360-degree feedback. [...]. I cannot have five different processes" (#9). On the other hand, employees argued for keeping it separate from the formal processes to keep it casual, fun and engaging as well as prevent dishonest use. Further observations include facilitating factors, e.g., when working "on-the-go or if you have gap time between meetings" (#1). It was emphasized that the app "should necessarily be seen together with other digitalization topics that we are talking about here at [organizational entity], for example digital e-learning" (#21). In such contexts, people's efforts need to be recognized and incentivized to bring projects forward.

Social Subsystem. Employees are part of a wider social subsystem. We find work practices that facilitate alternative ways to exchange feedback and inhibit the need to use the feedback app. For example, closing meetings with face-to-face feedback rounds was mentioned as "ending ritual of meetings" (#1). Thus, limiting actualization to situations in which additional feedback is provided afterwards (see Path 4 in Figure 2). Also, daily Scrum stand-up meetings offer an alternative to exchange timely feedback. Further, working frequently with coworkers over rich media is identified as inhibiting, while extended and international team structures are facilitators: "of course I use it a lot more when I'm in [inter-regional meeting]. […]. They come in, present, go out, and fly back to Paris, Milano, and so on […] then you write that together in the evening, on your way home, if you sit on the train" (#19). This is increasingly relevant, because „when we are developing into [the direction of] virtual teams with less rigid hierarchies and work with different teams across projects, we just need it" (#21). In addition to prevailing structures, the pilot project entails inhibiting and facilitating context factors. For example, collecting feedback and points without incentives and possibilities to redeem these points was mentioned as inhibiting factor together with limited institutional guidance when to use the app: "I wish to have more guidance on when to give feedback. […, and on] how to understand the feedback app vs 360/ multi rater vs other regular feedback" (#FR). Sharing identified use scenarios and best practices was perceived to mitigate this factor. Also, actor-related context factors of the pilot further inhibited the actualization. Many users where surrounded by non-registered coworkers, and hence, felt that there are too little pilot users: "as the group is small, it's hard to not be too repetitive and/or biased towards the group who participates" (#FR). Further, being surrounded by actors that prefer to exchange feedback personally limits its usefulness. While general feedback on sensitive and controversial topics was preferably exchanged in person, employees' preferences varied for concrete and operational-level feedback. Coworkers with a positive attitude towards the idea of facilitating feedback and digital and social media capabilities foster the actualization.

5 Discussion and Conclusions

5.1 Discussion of Implications

Implications for Theory on Digital Work, Social Software, and Feedback. We address calls for research on digital work [3, 11] by elaborating how employees in digital work environments still navigate back and forth between various digital and physical spaces to perform work and exchange feedback. Our results reveal that the feedback app, in contrast to traditional feedback systems, is immediately perceived as digital work tool for operational-level performance feedback. As such, there are similarities to alternate systems that offer similar affordances. Namely operational feedback may be exchanged in general purpose social software (e.g., email, Slack), task-specific systems with embedded feedback mechanisms [18], and through integrating enterprise systems in social software that facilitates social interactions and feedback exchange [50]. While the feedback app is perceived as enabling personal development and growth, this second-order affordance requires employees to use the feedback app as central hub to accumulate feedback in one place. Therefore, future research should investigate how performance feedback can be integrated across systems (e.g., feedback app and enterprise systems) and sources (e.g., computer-mediated and computer-generated). Further, the feedback app introduces a novel type of enterprise social software aside from general-purpose social software such as social networks. Prior research on social software for specific purposes is scarce, hence, we provide unique contextual insights on social software tailored to the exchange of performance feedback. These findings are equally relevant for performance management literature, since they respond to calls for research to better investigate informal day-to-day feedback [51] and to examine the use of technology in managing performance [8].

Implications for Affordance Theory. Existing research draws on socio-technical systems theory to elaborate affordance emergence and perception [40] as well as the actualization of organizational-level affordances [41]. We extend prior research [40, 41] by elaborating how socio-technical context factors affect the actualization process of individuals situated in organizational work environments. We contribute to existing actualization models [35, 37] since the identified factors provide a concrete explanation of how perceptions of expected value and effort are affected by the socio-technical context and why affordances may not be actualized even though they are perceived.

Implications for Practice. Practitioners that introduce feedback apps, should mitigate the identified constraints and inhibiting socio-technical context factors, while enhancing facilitating factors. Designers of feedback apps should consider the use scenarios and trajectories by supporting these practices and preventing the identified constraints as well as inhibiting context factors. For example, feedback apps should integrate well into software in which digital work is performed and should address the need of employees to switch between physical and digital spaces.

5.2 Conclusion and Limitations

Investigating the feedback app within a pilot project at InsurCorp enabled us to elaborate affordances and constraints perceived and actualized in a naturalistic case setting as well as to present facilitating and impeding socio-technical context factors.

Nevertheless, our results must be viewed in the light of its limitations. The qualitative and interpretive nature of this research prevents exhaustiveness. The selected pilot participants and interviewees possibly share behavioral and perceptional traits that may not be representative. Even though we minimized selection bias by considering employees from diverse operational entities, locations and job roles, they all belong to the same large financial services group. Finally, changes of the technical and social structures as well as data analyses in this domain are restricted by the works council. Future research in other organizational contexts and countries is needed.

References

1. Baker, A., Perreault, D., Reid, A., Blanchard, C.M.: Feedback and organizations: Feedback is good, feedback-friendly culture is better. Canadian Psychology. 54, 260 (2013).
2. Ashford, S.J., Tsui, A.S.: Self-regulation for Managerial Effectiveness: The Role of Active Feedback Seeking. Acad. Manage. J. 34, 251–280 (1991).
3. Richter, A., Heinrich, P., Stocker, A., Schwabe, G.: Digital Work Design. Business & Information Systems Engineering. (2018).
4. Riemer, K., Schellhammer, S., Meinert, M.: Collaboration in the Digital Age: How Technology Enables Individuals, Teams and Businesses. Springer (2018).
5. Stoeckli, E., Dremel, C., Uebernickel, F.: Exploring characteristics and transformational capabilities of InsurTech innovations to understand insurance value creation in a digital world. Electronic Markets. 28, 287–305 (2018).
6. Durward, D., Blohm, I., Leimeister, J.M.: Crowd Work. Business & Information Systems Engineering. 58, 281–286 (2016).
7. Buckingham, M., Goodall, A.: Reinventing Performance Management. Harv. Bus. Rev. 93, 40–50 (2015).

8. Levy, P.E., Tseng, S.T., Rosen, C.C., Lueke, S.B.: Performance Management: A Marriage between Practice and Science--Just Say "I do." In: Research in Personnel and Human Resources Management. pp. 155–213. Emerald Publishing Limited (2017).
9. Levy, P.E., Williams, J.R.: The Social Context of Performance Appraisal: A Review and Framework for the Future. J. Manage. 30, 881–905 (2004).
10. Bailey, D.E., Leonardi, P.M., Barley, S.R.: The lure of the virtual. Organization Science. 23, 1485–1504 (2012).
11. Mrass, V., Li, M.M., Peters, C.: Towards a Taxonomy of Digital Work. In: ECIS 2017 Proceedings. p. 2515 (2017).
12. Arnold, D., Butschek, S., Steffes, S., Müller, D.: Monitor-Digitalisierung am Arbeitsplatz: Aktuelle Ergebnisse einer Betriebs-und Beschäftigtenbefragung. ZEW-Gutachten und Forschungsberichte (2016).
13. Kuhnen, C.M., Tymula, A.: Feedback, Self-Esteem, and Performance in Organizations. Manage Sci. 58, 94–113 (2012).
14. Chen, J.V., Ross, W.H.: The managerial decision to implement electronic surveillance at work: A research framework. Int J Organ Anal. 13, 244–268 (2005).
15. Ang, S., Cummings, L.L., Straub, D.W., Earley, P.C.: The Effects of Information Technology and the Perceived Mood of the Feedback Giver on Feedback Seeking. Information Systems Research. 240 (1993).
16. Kluger, A.N., Denisi, A.: The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol. Bull. 119, (1996).
17. Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., Staake, T.: Overcoming salience bias: how real-time feedback fosters resource conservation. Manage Sci. 64, 1458–1476 (2016).
18. Jung, J.H., Schneider, C., Valacich, J.: Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments. Manage Sci. 56, 724–742 (2010).
19. Wang, H., Wang, J., Tang, Q.: A Review of Application of Affordance Theory in Information Systems. Journal of Service Science and Management. 11, 56--70 (2018).
20. Reinhardt, W., Schmidt, B., Sloep, P., others: Knowledge worker roles and actions—results of two empirical studies. Knowledge and Process Management. (2011).
21. Hattie, J., Timperley, H.: The Power of Feedback. Rev. Educ. Res. 77, 81–112 (2007).
22. Farr, J.L.: Informal performance feedback: Seeking and giving. In: Schuler, H., Farr, J.L., and Smith, M. (eds.) Personnel Selection and Assessment: Individual and Organizational Perspectives (1993).
23. Baker, N.: Employee feedback technologies in the human performance system. Human Resource Development International. 13, 477–485 (2010).
24. Cappelli, P., Tavis, A.: The Performance Management Revolution. Harv. Bus. Rev. 94, 58–67 (2016).
25. Aral, S., Walker, D.: Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. Manage. Sci. 57, 1623–1639 (2011).
26. Stoeckli, E., Uebernickel, F., Brenner, W.: Capturing Functional Affordances of Enterprise Social Software. In: AMCIS 2017 Proceedings (2017).
27. Richter, A., Riemer, K.: Malleable End-User Software. Bus Inf Syst Eng. 5, 195–197 (2013).
28. Leonardi, P.M.: When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. MIS Quarterly. 35, 147–167 (2011).

29. Koch, M.: CSCW and Enterprise 2.0-towards an integrated perspective. BLED 2008 Proceedings. (2008).
30. Leonardi, P.M.: Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How are They Related? Do We Need Them? Materiality and Organizing: Social Interaction in a Technological World. 25 (2012).
31. Orlikowski, W.J., Barley, S.R.: Technology and Institutions: What Can Research on Information Technology and Research on Organizations Learn from Each Other? MIS Quarterly. 25, 145–165 (2001).
32. Gibson, J.J.: The Theory of Affordances. In: and Bransford, Shaw, Robert (ed.) Perceiving, Acting and Knowing. pp. 67–82. Lawrence Erlbaum Associates (1977).
33. Volkoff, O., Strong, D.M.: Affordance theory and how to use it in IS research. Routledge Handbooks Online (2017).
34. Markus, M.L., Silver, M.S.: A foundation for the study of IT effects: A new look at DeSanctis and Poole's concepts of structural features and spirit. JAIS. 9, 609 (2008).
35. Bernhard, E., Recker, J.C., Burton-Jones, A.: Understanding the actualization of affordances: A study in the process modeling context. In: ICIS 2013 Proceedings (2013).
36. Orlikowski, W.J.: The Duality of Technology: Rethinking the Concept of Technology in Organizations. Organization Science. 3, 398–427 (1992).
37. Glowalla, P., Rosenkranz, C., Sunyaev, A.: Evolution of IT Use: A Case of Business Intelligence System Transition. Presented at the ICIS 2014 Proceedings (2014).
38. Jung, Y., Lyytinen, K.: Towards an ecological account of media choice: a case study on pluralistic reasoning while choosing email. Info Systems J. 24, 271–293 (2014).
39. Bygstad, B., Munkvold, B.E., Volkoff, O.: Identifying generative mechanisms through affordances: a framework for critical realist data analysis. J Inf Technol. 31, 83–96 (2016).
40. Seidel, S., Recker, J.C., vom Brocke, J.: Sensemaking and sustainable practicing: Functional affordances of information systems in green transformations. MIS Quarterly. 37, 1275–1299 (2013).
41. Dremel, C., Herterich, M., Wulf, J., Brocke, J.V.: Actualizing Big Data Analytics Affordances: A Revelatory Case Study. Information & Management. (2018).
42. Bostrom, R.P., Heinen, J.S.: MIS Problems and Failures: A Socio-Technical Perspective, Part II: The Application of Socio-Technical Theory. MIS Quarterly. 1, 11–28 (1977).
43. Leavitt, H.J.: Applied organizational change in industry: Structural, technological and humanistic approaches. Handbook of organizations. 1144, 1170 (1965).
44. Hester, A.J.: Socio-technical systems theory as a diagnostic tool for examining underutilization of wiki technology. The Learning Organization. 21, 48–68 (2014).
45. Eisenhardt, K.M.: Building Theories from Case Study Research. Acad. Manage. Rev. 14, 532 (1989).
46. Corbin, J.M., Strauss, A.: Grounded theory research: Procedures, canons, and evaluative criteria. Qual. Sociol. 13, 3–21 (1990).
47. Yin, R.K.: Case study research and applications: Design and methods. Sage (2017).
48. Schultze, U., Avital, M.: Designing interviews to generate rich data for information systems research. Information and Organization. 21, 1–16 (2011).
49. Klein, H.K., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly. 23, 67–93 (1999).
50. Stoeckli, E., Uebernickel, F., Brenner, W.: Exploring Affordances of Slack Integrations and Their Actualization Within Enterprises - Towards an Understanding of How Chatbots Create Value. In: Proceedings of the 51th HICSS (2018).
51. Ashford, S.J., Cummings, L.L.: Feedback as an individual resource: Personal strategies of creating information. Organ. Behav. Hum. Perform. 32, 370–398 (1983).

The Effect of Marker-less Augmented Reality on Task and Learning Performance

Peter Sommerauer, Oliver Müller, Leonard Maxim, and Niels Østman

IT University of Copenhagen, Department Business IT, Copenhagen, Denmark
{psom,olmy,leom,nies}@itu.dk

Abstract. Augmented Reality (AR) technologies have evolved rapidly over the last years, particularly with regard to user interfaces, input devices, and cameras used in mobile devices for object and gesture recognition. While early AR systems relied on pre-defined trigger images or QR code markers, modern AR applications leverage machine learning techniques to identify objects in their physical environments. So far, only few empirical studies have investigated AR's potential for supporting learning and task assistance using such marker-less AR. In order to address this research gap, we implemented an AR application (app) with the aim to analyze the effectiveness of marker-less AR applied in a mundane setting which can be used for on-the-job training and more formal educational settings. The results of our laboratory experiment show that while participants working with AR needed significantly more time to fulfill the given task, the participants who were supported by AR learned significantly more.

Keywords: Augmented Reality, Learning, Mobile Application

1 Introduction

Augmented Reality (AR) is known as a technology which augments the real environment with relevant digital information [3]. Such information can be superimposed on recognized objects using smartphones, tablets or AR goggles as user interfaces between the real and the virtual world. Additionally, AR allows a full 3D view of virtual objects and enables users to interact with them.

AR's potential has been shown in many use cases and in various settings, such as informal and formal learning environments, workplaces, museums and natural environments [2, 4, 5, 6, 10, 12, 21, 22, 24, 25]. In most settings which have been studied so far, trigger images or QR codes have been used for identifying objects in order to superimpose digital information on them [4]. Only few applications exist that use so-called marker-less AR [4]. Marker-less AR works in a way that the real environment itself and real objects therein are recognized by the app, which then augments digital information and adds functionality to the digitally enriched objects and environments [25], without any pre-defined trigger images or QR codes.

In this study we investigate the application and effectiveness of marker-less AR to support both the execution of a specific task in a mundane setting and the learning about

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the underlying domain by executing the task (i.e., learning-by-doing). In particular, we intend to answer the following research questions:

- RQ1: How can marker-less AR be implemented in a real-world environment?
- RQ2: How does marker-less AR affect task and learning performance?

In the pursuit of answering our research questions, we developed a marker-less AR app, which enables the user to learn the names of objects from the real environment. We created a fictional learning situation with a given task and compared the results from two groups, one using an AR-based tool, the other using a traditional paper-based tool (Note that a direct comparison between marker-less and marker-based AR is not the aim of this study). Hence, our laboratory experiment uses a static group design with an experimental group and a control group. With this design we intent to investigate the differences in task and learning performance of the two groups by measuring task performance (i.e., time required for completing the task) and learning performance (i.e., answering a post-test questionnaire with questions about the task).

The remainder of this paper is structured as follows: To prepare the background, we first present related work and provide theoretical background on marker-less AR and its implementation. As our study was motivated by investigating task performance and learning performance, we also present associated performance metrics that are derived from learning theories. We then outline the app development process along with the embedding of a number of theory-ingrained design principles, followed by an introduction of the used dataset for image recognition and the setup and execution of our experiment. Next, we provide detailed insights into our data analysis, which prepares for the discussion of our results. Finally, we conclude with a brief summary and directions for future research.

2 Background

Our research background focuses on synthesizing the findings of published systematic literature reviews on AR learning and empirical studies about marker-less AR from the last decade. In order to identify relevant related work, we analyzed the most cited literature reviews on AR for education.

Most extant studies do not focus on using AR in real-life environments, but investigate its use for supporting a narrow and well-defined task in a controlled setting. Hence, it is not surprising that virtually all existing studies focus on the application of marker-based AR, which is easy to implement in a controlled laboratory setting, and that only few studies have investigated the use of marker-less AR so far [4, 5, 21, 26]. Moreover, marker-less AR is one key aspect discussed for implementing hybrid tracking for ubiquitous AR [5, 21, 25, 26].

What most studies have also in common is that they emphasize the need of further research on the features, use, advantages, and limitations of AR in educational settings [2, 4, 6, 10]. Reported advantages of AR in educational settings include learning gains, higher motivation, facilitated interaction, better collaboration, lower cost, better user experiences, just-in-time information, enabling of situated learning and student-centered approaches, increase of students' attention, enjoyment, exploration, increased capacity for innovation, creation of positive attitudes, more awareness, anticipation, and authenticity [2, 3, 4, 5, 6, 10, 11, 12, 21, 22, 24, 25, 26]. In contrast, repeatedly reported limitations of AR in education include the observation that AR apps are mostly designed for only one specific knowledge field [4], that teachers cannot create new learning content [2, 4, 10, 22], that there are difficulties maintaining superimposed information, that learners pay too much attention to the virtual information, that evaluation focused on short-term instead of long-term learning [4], and that AR can be perceived as an intrusive technology [4, 21, 26]. Still, most studies found positive evidence for the effectiveness of AR in education, for example, in the form of enhanced learning performance, higher learning motivation, improved perceived enjoyment, decreased cost, as well as adding creating positive attitudes towards education and fostering students' commitment [2, 3, 4, 5, 6, 10, 11, 12, 21, 22, 24, 25, 26].

In Bacca et al.'s review of AR for education, the authors report about 19 studies that use marker-based AR, 4 studies with marker-less AR, and 7 studies covering location-based AR [4]. They discuss challenges around the improvement of recognition algorithms (e.g., for human forms) in the process of achieving more immersive and not intrusive AR learning experiences. Furthermore, they recommend vocational educational training (VET) classes as target groups for future studies.

In their literature survey of AR, Billinghurst et al. [5] additionally focus on technology for user activity tracking considering input and interaction. They provided first design guidelines and interface patterns for AR development tools, starting with considering physical objects, virtual content and interaction metaphors and their connection. Additionally, they suggest future research directions as user tracking, user interaction, AR displays, and social acceptance of AR.

Dunleavy & Dede provide insights in AR teaching and learning, focusing on AR utilizing mobile, context-aware technologies (e.g. smartphones, tablets), thus enabling AR users interacting with digital information which is embedded within physical environments and in both, formal and informal learning environments [12]. They additionally investigate affordances and limitations for AR related to teaching, learning and instructional design and see AR as primarily aligned with situated and constructivist learning theory, stating, that AR positions learners within a real-world physical and social context while guiding, scaffolding and facilitating participatory and metacognitive learning processes (e.g. authentic inquiry, active observation, peer coaching, reciprocal teaching). Since AR legitimate users in peripheral participation with multiple modes of representation, they distinguish between location-aware and vision-based AR. In this context, AR has some limitations regarding student cognitive overload and managing level of complexity, which is a key instructional issue. Therefore, they recommend to decrease cognitive load by creating a simplified experience structure initially and increasing complexity as the experience progresses, thus scaffolding each experience explicitly at every step to achieve the desired experience or learning.

When Radu states that the educational community remains unclear regarding the educational usefulness of AR and regarding contexts in which this technology is more effective than other educational mediums, he refers to 26 publications comparing student learning with AR vs. non-AR apps [22]. Radu observed some negative consequences, such as attention tunneling, usability difficulties, ineffective classroom integration, and learner differences. His table of factors influencing learning in AR covers content representation, multiple representations that appear at appropriate time and space, learners are physically enacting educational concepts, attention is directed to relevant content, learners are interacting with 3D simulations, interaction and collaboration are natural.

Still, the benefits of AR in educational environments and the value of AR apps applied in educational environments has not yet been investigated in its entirety [8]. The different directions of AR apps differ regarding their potential benefits. In their systematic literature review to synthesize a set of 25 publications, Diegmann et al. [10] identified 14 different benefits clustered in six different groups. They considered dimensions like state of mind (e.g. increased motivation, increased attention, increased concentration, increased satisfaction), teaching concepts (e.g. student-centered learning, collaborative learning), presentation (e.g. increased details, information accessibility, interactivity), learning type (e.g. improved learning curve, increased creativity), content understanding (e.g. improved development of spatial abilities, memory), and reduction of costs [10]. They then mapped the benefits to five directions of AR in educational environments (discovery-based learning, objects modeling, AR books, skills training, AR gaming) and indicated that specific directions of AR apps are more likely to lead to certain benefits, such as increased motivation. Especially, they emphasize that future research is needed to investigate the causality between benefits and directions of AR.

In their review of AR in education from 2011 to 2016, Chen et al. focused on research which includes the uses, advantages, features, and effectiveness of AR in educational settings [6]. They recommended to undertake more studies considering the difference of cognitive process and psychological immersion between AR and reality settings, individual interaction, sense of identity, adaptive application in AR, AR classroom design and evaluation research, teacher's role model in AR educational setting, design and implementation of AR learning resources in K-12.

The literature review by Akçayır & G. Akçayır focuses on current advantages and challenges of? AR education. Although AR promotes enhanced learning achievement, they experienced a discrepancy for AR in terms of cognitive load and/or cognitive overload, and AR ease of use vs. challenges for AR app usability [2]. Since research studies report both, they advise AR developers to develop and consequently implement empirically proven design principles, focusing on AR use and educational outcomes, and AR apps designed for diverse populations (e.g. kids, students, lifelong learners). They emphasize the need to investigate students' satisfaction, motivation, interaction, and commitment, and provide insights from research and development comprising explanations of development processes and factors being considered in design.

Dunleavy, Dede, and Mitchell document in their review covering AR simulations for teaching and learning, how teachers and students describe and comprehend ways of participation in AR simulation, to aid or hinder teaching and learning [11]. By means of qualitative case studies across two middle schools they demonstrate that AR supports multi user environments and immersive collaborative simulation.

For professional education and training, Palmarini et al. focused on the state of the art of AR apps applied in maintenance [21]. Based on 30 primary studies between 1997-2017, they unveil most relevant technical limitations for AR and propose results indicating a high fragmentation among hardware, software and AR solutions which lead to a high complexity for selecting and developing AR systems, thus identifying areas where AR technology still lacks maturity (e.g. marker-less AR).

Further limitations for AR which are still present today were depicted by Zhou et al. for tracking techniques, interaction techniques, user interfaces, and AR displays, especially for head mounted displays (HMD) [26]. Although the development of AR hardware became more sophisticated in the past decade, the major technical issues are not sufficiently dissolved and need to be overcome, like low sensitivity trigger to recognition [2].

3 Methodology

In our study we followed the advice from Dunleavy & Dede (i.e., decreasing cognitive load by creating a simplified experience structure) [12], Diegmann et al. (i.e., causality between benefits of AR) [10], Chen et al. (i.e., AR classroom design and evaluation research, design and implementation of AR learning resources) [6], Palmarini (i.e., use of marker-less AR) [21]. In order to develop an AR app for both school and professional education (VET), we applied design principles from Billinghurst et al. (i.e., real physical objects/virtual elements to be displayed, linking interaction metaphor) [5] and Sommerauer & Müller (i.e., design elements derived from learning theories) [24].

For the evaluation of the effect of marker-less AR applied in a learning scenario we chose to design a controlled laboratory experiment to compare the support of AR with traditional, paper-based material inside a classroom. In this, we aimed to ensure that no or hardly any differences in information equivalence [16] could affect the results of our study. Finally, our research design aimed to support and control exactly those research design elements which were the key subject of investigation.

With the experiment we examined the usability of AR, its effectiveness and the potential for teaching and learning. The evaluation covered measures for perceived usefulness, perceived learning and students' motivation as well as objective performance in terms of time to completion for the task and number of mistakes made in a recall and retention test administered as a post-test. In addition, we employed the Systems Usability Scale (SUS) [23] to evaluate the usability of the applied AR system.

In our app development, we considered design elements from Billinghurst et al., who proposed to focus on physical objects, virtual content, the interaction metaphor, and their connections [5]. Additionally, we applied the conceptual framework by Sommerauer & Müller [24], which is inspired by Anderson's work on how learning can be enhanced using emerging technologies and applying learning theories [1]. At the heart of this framework are one or more learning sequences, each consisting of one or more connected learning activities. At the center of a single learning activity stands the learning content. This content should be designed according to different learning theories, indicated by the different concentric layers surrounding the learning content.

At the first layer, it is proposed to apply the 12 design principles of the cognitive theory of multimedia learning (CTML) [18]. In the second layer, design elements from mobile learning (e.g., Herrington et al. [13]) shall be considered for application design. Finally, it is proposed to implement design elements from game-based learning (e.g. leaderboard, mission) [14], simulations (e.g., storytelling, drama), experiential learning theory (e.g., diverging, assimilating) [15], and situated learning [19]. Additionally, collaborative learning elements can be introduced at the learning stage, where multiple learning activities are combined into a learning sequence [24].

We instantiated the above described conceptual framework by developing an AR learning app prototype. It supports the task of learning names related to physical objects used in a particular professional domain – in our case, the florist industry. More specifically, the app combines machine learning techniques for image recognition and machine translation to identify objects that are in the focus of the mobile phone camera in real-time and superimpose information such as the object's name in different languages onto the object. As a training application, the app can be used in any workplace environment and the trainee can select between exploration mode or quiz mode. In both, the user needs to focus the particular object using the device's camera (e.g. smartphone, tablet, any head-mounted device). Once the object is recognized, the app provides a selection of labels, comprising the three most likely names of the object using a percentage scale and colors. In quiz-mode, the app shows the most likely label and two randomly selected labels and the trainee has to pick the correct one. Figure 1 shows screenshots of the application and show the explore and quiz modes.

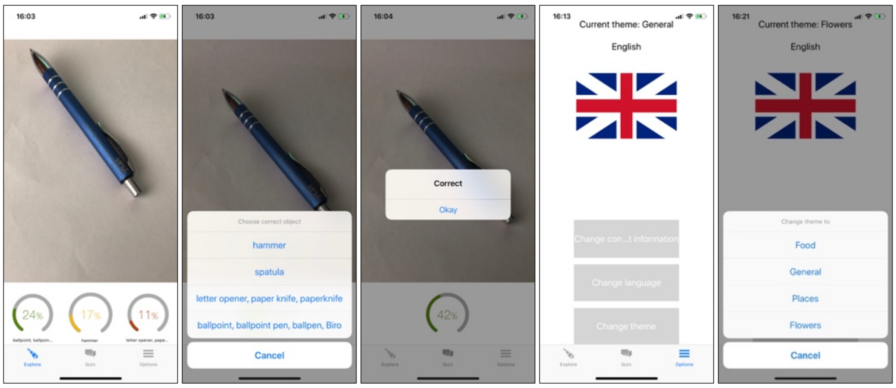


Figure 1. App in explore mode, quiz mode and selection of language and theme

The app design integrates design elements from CTML (i.e., the multimedia principle, the spatial contiguity principle, the temporal contiguity principle, and the signaling principle) with elements from the theory of mobile learning (i.e., users can use the app across space and time) and game-based elements). From a technical perspective, the app is based on Apple's ARKit framework[1] for implementing mobile

[1] https://developer.apple.com/arkit/

AR experiences, Google's MobileNets model[2], a convolutional neural network for efficient image recognition on mobile phones, and the Google Translate API[3] for automated translation of texts into multiple languages.

As a foundation for our flower identification app we used the flowers dataset by Nilsback & Zisserman [20], implemented as a selectable theme in our app. The flowers dataset consists of 8,189 images of flowers commonly occurring in the United Kingdom. The images are divided into 103 classes and each class consists of between 40 and 250 images. The images are scaled so that the smallest dimension is 500 pixels. The flowers are identified by different features describing different properties, e.g., color (HSV values of pixels), histogram of gradient orientations (HOG) [8], and distinctive image features (SIFT) [17] on foreground region and foreground boundary. In prior studies the recognition accuracy was measures at 72.8 percent.

The instructional design for the learning situation applied in the experiment contained elements from cognitive and constructivist learning theories. While the learning content was prepared based on CTML principles, elements of constructivist theory were implemented in the learning activity, such as, task orientation, mobile learning, and situated learning, by sending learners on missions including storytelling.

The laboratory experiment was based on a sequential quantitative method research [7] applying a static group design. The aim of the experiment was to identify differences in the application of AR vs. traditional learning. While the experiment group was supplied with mobile devices (iPhone 8+ and X) running the AR app, the control group received a traditional, paper-based tool (catalogue) to fulfil their task. Both groups received the same instructions and were required to fulfil the same task. At the end of the experiment both groups received a post-test questionnaire covering the same topics and questions. The questionnaire contained three sections. The first covered aspects for perceived usefulness, perceived learning, and students' motivation. The second part was a multiple-choice test asking for the names of five flowers shown as pictures. For each, participants could choose between three given names. The number of correctly identified flowers was used as an objective measure for learning performance. The third section of the questionnaire contained ten questions from the System Usability Scale (SUS), which was only available for the group using AR in the experiment.

4 Experimental Setup

The laboratory experiment followed a static group design comprising an experimental group and a control group. With this design we intended to investigate the differences in task and learning performance of two groups: one supported by an AR tool and one using traditional tools (i.e. a catalogue). Following similar studies [2, 3, 4, 5, 6, 11, 12, 22, 24] and in line with our research questions, we used the item "time for task completion" as a measure for task performance and "No. of correctly identified flowers" from the questionnaire after the treatment as an indicator for learning performance. Figure 2 gives an overview of the randomized field experiment.

[2] https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet

[3] https://cloud.google.com/translate/docs/apis

			Treatment				Measurement					
				collect flowers	apply tool		time	performance	perceived usefulnes	perceived learning	learning	motivation
Instructions for participants: welcome, acknowledgement, rules for the experiment, motivational frame story	**Random Assignment to AR group/Non-AR group**	**Group 1**	select one out of five envelopes: includes task description and randomized collection of 6 flower names	collect 6 flowers	smartphone with AR app	fill in questionnaire	time to task completion	No. of correct selected flowers	one question in questionnaire	3 questions in questionnaire	multiple choice test presenting flowers: tick the correct name	4 questions in questionnaire
		Group 2			paper based catalogue							

Figure 2. Overview of the randomized field experiment

We prepared two flower meadow, each consisting of 100 fake flowers composed of four different flower pictures per flower species and covering a selection of 25 different flower species from the flower dataset. The pictures were printed on paper and mounted on skewers. On the back side, the fake flowers were numbered according to an internal reference list to allow internal identification without the need for labels.

As a traditional tool for supporting participants in the experiment, we prepared a flowers catalogue covering exactly the 25 different flower species from the flower meadows. The flower pictures in the catalogue were different from those in the flower meadow and the catalogue was ordered alphabetically.

The questionnaire in the first section used a Likert scale containing five values from strongly disagree (1) to strongly agree (5) and covering eight questions:

- Perceived Usefulness:
  A. The AR app / catalogue was helpful to fulfil the task.
- Perceived Learning:
  B. With this activity I have learned something.
  C. I have learned about flowers.
  D. I can put together a bouquet on my own.
- Motivation: What do you think about the experiment and its setup?
  E. The introductory story was motivating.
  F. The task was simple and understandable.
  G. It was exciting to fulfill the task.
  H. The activity was entertaining.

Both rooms for the experiment where prepared in the same way. We set up the flower meadow with the fake flowers sticking in carton boxes and grouped by flower type. The carton boxes were placed on three tables in the center of the room. There was enough space to walk around the tables and to reach the flowers easily.

The main task for the participants was to collect six flowers from the meadow, which were named in form of a word-cloud on the instruction sheet in an envelope. We prepared five envelopes and the selection of the flower names for the word-cloud was done by a randomization process. Such, we used a webtool (www.randomizer.org) to collect 5 sets of 6 unique numbers per set within the range from 1 to 25. To arouse student attention and motivation, we narrated a story to send them on a mission, thus following design principles from game-based learning and simulation: “You fell in love with another person and have learned that you can break the ice between you and your crush with a smoothly arranged bouquet of flowers. Since you are absolutely unfamiliar with how to create a convincing flower bouquet, you ran a data analysis on your partner’s Facebook account and received a list of preferred flowers presented in the word-cloud below”. The mission to accomplish was formulated in the way, that “You know that love is like a little bird which flies away after some time and since you have just this one chance to score, give your best and collect the flowers as listed in the word cloud from the “self-service shop” as accurately and as fast as you can!”.

While the AR group could use a prepared iPhone (we used four iPhone 8+ and one iPhone X) to complete their mission, the control group (non-AR group) was provided with the aforementioned flowers catalogue. As noted earlier, we used different pictures for the catalogue and the production of the fake flowers.

The experimental process was designed in a way that after listening to the initial instruction participants were assigned an envelope with further instructions, the story, the mission, and either an iPhone or a flowers catalogue. Then the researcher started a timer and the students needed to collect the flowers as fast as possible. Afterwards, they came back to the researcher who recorded the collected flower numbers and asked students to complete the questionnaire. Since the students received a participant number, this number was noted on the questionnaire for later analysis. Once the participants completed all tasks, the fake flowers were put back to the flower meadows and the room was prepared for the next group.

5 Implementation

We invited 71 students from a Masters course in Information Technology at a technical university in northern Europe to participate in the experiment, but only 44 attended. The students were already divided into working groups from their course and we assigned them to sessions with a maximum of ten students per session and a duration of approximately 15 minutes. Participating students received a voucher from the university’s coffee shop as a reward right after the experiment.

The experiment started with a short introduction to welcome and thank the students for their participation. The participants were given some motivational instructions and were told to not chat with each other during the experiment or tell others about the experiment afterwards to not influence other students attending later. To split the group into the AR group (participants interacting with AR app during the experiment) and non-AR group (control group working with catalogue instead of AR app), students were told to choose between one of the two rooms by having equal numbered groups.

Participants could choose one of the five envelopes and when they started reading the instructions, a timer was set. After collecting the flowers, the students had to move to the research assistant and hand over their flower bouquet and all provided materials. To document the selected flowers and the required time to completion for the task, participants received a number to record their results for analysis. They then received the questionnaire to be answered on their own, marked with their participants number. After the students completed the questionnaire, they could leave the experiment.

Both experiment groups were treated in the same way, except of having different tools (AR app and paper catalogue) to fulfill the main task. There were no a priori time restrictions given, but students in the AR group were asked to terminate the collecting of flowers after 15 minutes.

6 Data Analysis

A participants' data record contained participant ID, group (AR, non-AR), gender (female, male), envelope number, IDs of the collected flowers, time to task completion, and the answers to the questions of the post-test questionnaire. In a first analysis, we assessed the number of correct flowers collected and the answers from the questionnaire. Overall, 18 female and 27 male students took part in the experiment, where 20 were assigned to the AR group and 25 the non-AR group.

In the AR group, 6 female and 14 male participants required from 510 to 1200 seconds to complete the given task (median 858.5 seconds, mean 864 seconds). They collected between 2 and 6 correct flowers from the given bouquet (median 4, mean 4.45). In terms of learning performance, the number of correct named flowers in their post-test questionnaire reached from 0 to 5 (median 2, mean 2.55).

Table 1. Correlation matrix

		Group	Gender	Envelope	t2compl	NoCorrFl	QuizRes	QA	QB	QC	QD	QE	QF	QG	QH
Group	Pearson Correlation	1	-,183	,113	-,927**	,287	-,229	,382**	-,241	-,352*	-,079	-,026	,099	-,014	,117
	Sig. (2-tailed)		,230	,458	,000	,056	,130	,010	,111	,018	,607	,866	,519	,930	,444
Gender	Pearson Correlation	-,183	1	-,220	,139	-,325*	-,283	-,363*	,154	,149	,150	,000	-,011	-,115	-,158
	Sig. (2-tailed)	,230		,146	,364	,029	,059	,014	,312	,329	,325	1,000	,942	,452	,300
Envelope	Pearson Correlation	,113	-,220	1	-,053	,122	,099	-,001	,065	-,098	,058	,231	-,091	,241	,235
	Sig. (2-tailed)	,458	,146		,732	,425	,516	,994	,670	,522	,703	,126	,552	,110	,121
t2compl	Pearson Correlation	-,927**	,139	-,053	1	-,306*	,295*	-,326*	,242	,324*	,120	,052	-,034	-,009	-,172
	Sig. (2-tailed)	,000	,364	,732		,041	,049	,029	,109	,030	,433	,733	,824	,951	,259
NoCorrFl	Pearson Correlation	,287	-,325*	,122	-,306*	1	,107	,176	-,227	-,281	-,053	-,037	-,013	,232	,201
	Sig. (2-tailed)	,056	,029	,425	,041		,485	,249	,133	,061	,730	,812	,933	,126	,186
QuizRes	Pearson Correlation	-,229	-,283	,099	,295*	,107	1	,166	-,248	-,119	,021	,230	,196	,183	,199
	Sig. (2-tailed)	,130	,059	,516	,049	,485		,276	,101	,435	,890	,128	,197	,228	,189
QA	Pearson Correlation	,382**	-,363*	-,001	-,326*	,176	,166	1	-,188	-,069	-,290	,257	,354*	,306*	,328*
	Sig. (2-tailed)	,010	,014	,994	,029	,249	,276		,216	,652	,053	,089	,017	,041	,028
QB	Pearson Correlation	-,241	,154	,065	,242	-,227	-,248	-,188	1	,709**	,428**	,109	,034	,241	,148
	Sig. (2-tailed)	,111	,312	,670	,109	,133	,101	,216		,000	,003	,476	,824	,111	,331
QC	Pearson Correlation	-,352*	,149	-,098	,324*	-,281	-,119	-,069	,709**	1	,322*	,148	-,031	,274	,238
	Sig. (2-tailed)	,018	,329	,522	,030	,061	,435	,652	,000		,031	,331	,842	,068	,116
QD	Pearson Correlation	-,079	,150	,058	,120	-,053	,021	-,290	,428**	,322*	1	,018	-,173	,109	,060
	Sig. (2-tailed)	,607	,325	,703	,433	,730	,890	,053	,003	,031		,908	,256	,475	,696
QE	Pearson Correlation	-,026	,000	,231	,052	-,037	,230	,257	,109	,148	,018	1	,175	,506**	,545**
	Sig. (2-tailed)	,866	1,000	,126	,733	,812	,128	,089	,476	,331	,908		,250	,000	,000
QF	Pearson Correlation	,099	-,011	-,091	-,034	-,013	,196	,354*	,034	-,031	-,173	,175	1	,221	,127
	Sig. (2-tailed)	,519	,942	,552	,824	,933	,197	,017	,824	,842	,256	,250		,144	,407
QG	Pearson Correlation	-,014	-,115	,241	-,009	,232	,183	,306*	,241	,274	,109	,506**	,221	1	,674**
	Sig. (2-tailed)	,930	,452	,110	,951	,126	,228	,041	,111	,068	,475	,000	,144		,000
QH	Pearson Correlation	,117	-,158	,235	-,172	,201	,199	,328*	,148	,238	,060	,545**	,127	,674**	1
	Sig. (2-tailed)	,444	,300	,121	,259	,186	,189	,028	,331	,116	,696	,000	,407	,000	

In the non-AR group, 12 female and 13 male participants needed between 68 and 330 seconds to complete the task (median 171 seconds, mean 182.24 seconds). They collected 0 to 6 correct flowers from the given bouquet (median 5, mean 5.16) and the number of correctly named flowers in the post-test questionnaire reached from 0 to 5 (median 2, mean 1.96). Between the two groups there was no difference in the distribution of envelopes, which was tested by performing a Kolmogorov Smirnov test.

Next, we ran an exploratory correlation analysis between all relevant pairs of variables in our dataset (Table 1). We found statistically significant correlations between group assignment and time to completion (mean of AR/non-AR: 864sec/182sec), perceived usefulness (QA) (mean of AR/non-AR: 3.65/4.40), and one of the questions related to perceived learning (QC) (mean of AR/non-AR: 3.4/2.8). Interestingly, we also found a significant correlation between gender and the number of correctly collected flowers (mean of female/male: 5.33/4.52, $p<0.01$), and perceived usefulness (mean of female/male: 4.50/3.77, $p<0.01$).

As our pseudo random assignment of students to groups did not produce an even distribution of males and females between the AR and non-AR group and because the correlation analysis indicated that gender is correlated with some of our dependent variables of interest, we decided to use regression models to test the main hypotheses of our experiment, namely that AR has a positive impact on (perceived) task performance and (perceived) learning performance. The advantage of a regression model over t-tests or ANOVA is in the ability to model the influence of multiple independent variables (in our case group and gender) on one dependent variable. Table 2 summarizes the results of this analysis.

Table 2. Regression results

	Dependent variable:				
	Correct Flowers	Time to Completion	Perceived Usefulness	Questions Correct	Perceived Learn
	(1)	(2)	(3)	(4)	(5)
tercept	5.524***	188.992***	4.690***	2.381***	2.542***
	(0.303)	(33.946)	(0.222)	(0.301)	(0.278)
roup [AR]	-0.571	702.524***	-0.571*	0.857*	0.540
	(0.364)	(40.735)	(0.266)	(0.362)	(0.334)
ender [Male]	-0.700	-12.985	-0.559*	-0.810*	0.316
	(0.367)	(41.039)	(0.268)	(0.364)	(0.336)
bservations	44	44	44	44	44
2	0.151	0.881	0.208	0.181	0.091
djusted R^2	0.110	0.875	0.169	0.141	0.047
esidual Std. Error (df = 41)	1.179	131.997	0.862	1.172	1.081
Statistic f = 2; 41)	3.648*	152.021***	5.369**	4.531*	2.059

$^{*}p<0.05$; $^{**}p<0.01$; ***p<0.(

According to the regression results, participants in the AR group did not perform significantly better in terms of correctly identifying flowers than participants in the paper catalogue group. With regard to time needed to complete the task, participants in AR group even performed significantly worse than participants in the paper catalogue group. Hence, we did not find any empirical evidence that the AR app increased participants' objective task performance in terms of task accuracy and task time. Consistent with this finding, participants in the AR group evaluated the perceived usefulness of their tool (i.e. the AR app) significantly worse than participants in the non-AR group working with the paper catalogue.

However, when looking at objective learning performance, measured by the number of questions answered correctly in the post-test questionnaire, we found that participants in the AR group performed significantly better. This finding provides empirical support for the effectiveness of AR as a tool to enhance students' objective learning performance. With regard to perceived learning (measured by the average scores of questions B-D), we did not find a significant difference between the groups.

7 Discussion

In our experiment students achieved an observably better learning performance when using the AR flower identification app instead of a comparable paper catalogue, a result that is similar to prior research results comparing AR-based training to traditional paper-based training methods [4, 12, 22]. Therefore, and to answer RQ2, we conclude that AR can support students' learning performance. However, it may also be that the learning performance for the AR group was influenced by their longer task completion times, thus students were more engaged with the learning content and more motivated [2]. This can either be seen as a potential confounding factor which has to be controlled for in future studies (e.g. by predefining the available time for conducting a task), or as a positive side effect of using AR for teaching and learning [4, 5, 6, 12, 24, 25]. One could argue that when using AR students voluntarily spend more time with the learning materials, as compared to using traditional paper-based tools.

Considering participants' behavior during the experimental task, we noted that students in the AR group acted differently than those in the non-AR group. While participants in the AR group needed to investigate the flowers sequentially (because the app can only identify one object at a time) and thus examined nearly all flowers from the meadow, participants in the non-AR group selected a flower's name from the task description, searched for the name in the catalogue, and then located the flower by scanning the flower meadow with their eyes and matching the picture from the catalogue with the pictures on the meadow. On the one hand, this resulted in much shorter task times, as the human eye can focus on multiple objects at the same time (or at least can change focus much more quickly than AR technology), in comparison to the participants in the AR group who additionally had to perform the task of hand-eye coordination when using the app. On the other hand, when filling out the post-test questionnaire students realized that they had not inspected all flowers from the meadow and catalogue in sufficient detail in order to answer the questions correctly (the flowers students had to name in the post-test were different from those they had to collect).

A further observation related to the above point was that as participants in the AR group were forced by the app's functionality to look at each flower and since the app showed the three most likely names for identifying a flower and the elated confidence levels, students required more attempts to select the correct flower. We are convinced that this was a main driver behind the longer time needed to complete the task. Additionally, students from the AR group confirmed that it is more fun to look at the flowers with the app instead of just learning from a book.

It is remarkable that while the perceived learning of the AR group is not significantly higher compared to the non-AR group, their objective learning performance was significantly higher. The better objective learning performance may be explained by the different ways participants approached the task in the two groups. While students in the non-AR group focused on finding the flower picture for the given flower name and selecting a similar flower from the meadow, students in the AR group pointed their smartphone upon every single flower in the meadow to see its name. A single flower was represented multiple times in the flower meadow and students from the AR group visualized a particular flower more often. This finding corresponds to results from other studies, where AR is more effective than using traditional media [2, 4, 6, 10, 22, 24].

Since the paper catalogue prepared for the experiment was ordered alphabetically and only contained few pages covering the presented 25 flowers, students in the non-AR group had an advantage when matching flower names between the task description and catalogue. This could be a major limitation in our study in regard to the results for participants task performance times compared with participants from the AR group. Using a flower identification book with hundreds of pages ordered by species instead of alphabetically would have been more realistic for our comparison and would probably have led to different results, at least in terms of task completion times. However, this observation indicates that the prepared catalogue was designed to support task completion.

Our app is technically able to identify up to 60 pictures per second, comparing it with several thousands of pictures from the database. Thus, the setup of the experiment with only a handful of flowers did not challenge the full potential of the app, which is a further limitation in terms of system performance in comparison of traditional tools with AR based tools. Nonetheless, with our study we could contribute to the discussion about improvement of AR recognition and marker-less AR [4]. For future research and practical application, the AR app can be utilized in any other learning environment just by exchanging the underlying image recognition machine learning model. This represents a cost-efficient alternative to integrate AR into classroom trainings [22].

8 Conclusion

With the app development and its application in the experiment we could answer our RQ 1 and demonstrate how marker-less AR can be implemented for education in a real-world environment. Thus, we followed recommendations for further research in the directions of implementing AR in real-life settings [6] and applying image-based tracking [5] and marker-less AR [4, 21, 26] for ubiquitous learning [2]. Moreover, with our study we investigated how marker-less AR affects task and learning performance in a mundane setting, for example in our simulation of a florist’s job. Our results showed that from a learning aspect, students using the AR app performed better when it comes to recalling the learning content, similar to prior studies [2, 6, 10, 24]. Although students in both groups achieved the same level of accuracy in fulfilling the given task, those students in the AR group needed more time. Since the experimental setup unintentionally supported the control group in faster task completion time, which points towards the finding that tasks processed with AR need to be designed differently.

Relying on the predefined dataset and machine learning model from Nilsback & Zisserman [20] was an efficient decision and guaranteed a consistent recognition rate for each object in the experiment. However, participants had some troubles with finding the correct focus for the fake flowers because of reflections, shadows and different illumination caused by the changing daylight which is also mentioned in prior studies and therefore a limitation which should be investigated in future research [2, 11, 21].

Students from the non-AR group benefited from the reduced catalogue to accomplish their mission. Since the AR app is able to recognize up to 60 pictures in a second from a dataset containing 8,189 pictures, the comparison of both tools in the experiment and for the given task, to search and identify a flower by its given name, was not really fair. However, with our study we demonstrate a content application of AR in association with its benefits and directions, particularly its scalability in a mundane situation.

Conducting an experiment just with students is not always satisfying. However, in our larger research program this was only a first test to demonstrate the use of the marker-less AR app and to collect and analyze first empirical data to investigate its effectiveness. In fact, we are beyond this now and are currently testing the app with a target group of low-threshold skilled employees.

From the aspect of using marker-less AR in educational settings we have ascertained that the recognition sometimes lacks due to optical influences, which is still a common issue for AR applications [2, 4, 5, 12, 21, 24, 25, 26]. Hence, future technological development should focus on recognition algorithms and the preparation of large and validated datasets in order to support the implementation of marker-less AR in education and in various real-life situations. Furthermore, the application of object detection instead of image recognition inside AR applications provides potential for new findings about how full 3D support for such AR apps assists learning and a better understanding. First results from our continuing research already confirm that object detection facilitates the recognition of a series of objects in one single viewpoint.

References

1. Anderson, T.: Theories for learning with emerging technologies. Emerging technologies in distance education (2016)
2. Akçayır, M., Akçayır, G.: Advantages and challenges associated with augmented reality for education: A systematic review of the literature. In: Educational Research Review, vol. 20, pp. 1-11 (2017)
3. Azuma, R. T.: A survey of augmented reality. Teleoperators & Virtual Environments, vol. 6(4), pp. 355-385 (1997)
4. Bacca, J., Baldiris, S., Fabregat, R., Graf, S.: Augmented reality trends in education: a systematic review of research and applications. In: Journal of Educational Technology & Society, vol. 17(4), pp. 133 (2014)

5. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. In: Foundations and Trends® in Human–Computer Interaction, vol. 8(2-3), pp. 73-272 (2015)
6. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using Augmented Reality in Education from 2011 to 2016. In: Innovations in Smart Learning, pp. 13-18. Springer, Singapore, (2017)
7. Creswell, J. W.: Research design: Qualitative & quantitative approaches. Sage Publications, Inc. (1994)
8. Dalal, N., Trigg, B.: Histogram of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886-893. IEEE Computer Society (2005)
9. DeLone, W. H., McLean, E. R.: Information systems success: The quest for the dependent variable. In: Information systems research, vol. 3(1), pp. 60-95 (1992)
10. Diegmann, P., Schmidt-Kraepelin, M., Van den Eynden, S., Basten, D.: Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. In: Wirtschaftsinformatik, vol. 3(6), pp. 1542-1556 (2015)
11. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. In: Journal of science Education and Technology, vol. 18(1), 7-22 (2009)
12. Dunleavy, M., Dede, C.: Augmented reality teaching and learning. In: Handbook of research on educational communications and technology, pp. 735-745. Springer, New York (2014)
13. Herrington, A., Herrington, J., Mantei, J.: Design principles for mobile learning (2009)
14. Hirumi, A., Appelman, B., Rieber, L., Van Eck, R.: Preparing instructional designers for game-based learning: Part 2. In: TechTrends, vol. 54(4), pp. 19 (2010)
15. Kolb, D. A.: Experiential learning: Experience as the source of learning and development. FT press (2014)
16. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. In: Cognitive Science, vol. 11 (1), 65–100 (1987)
17. Lowe, D.: Distinctive image features from scale-invariant Keypoints. In: International Journal of Computer Vision, vol. 60(2), pp. 91–110 (2004)
18. Mayer, R. E.: Multimedia learning (2nd). Cambridge University Press (2009)
19. McLellan, H.: Evaluation in a situated learning environment. In: Situated Learning perspectives, Englewood Cliffs, NJ: Educational Technology Publications (1996)
20. Nilsback, M. E., Zisserman, A.: Automated flower classification over a large number of classes. In: Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing (2008)
21. Palmarini, R., Erkoyuncu, J. A., Roy, R., Torabmostaedi, H.: A systematic review of augmented reality applications in maintenance. In: Robotics and Computer-Integrated Manufacturing, vol. 49, pp. 215-228 (2018)
22. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. In: Personal and Ubiquitous Computing, vol. 18(6), pp. 1533-1543 (2014)
23. Sauro, J.: A practical guide to the system usability scale: Background, benchmarks & best practices. Denver, CO: Measuring Usability LLC (2011)
24. Sommerauer, P., Mueller, O.: Augmented Reality for Teaching and Learning – a Literature Review on Theoretical and Empirical Foundations. In: Proceedings of the 26th European Conference on Information Systems (ECIS), Portsmouth, UK, June 23-28 (2018)
25. Van Krevelen, D. W., Poelman, R.: A survey of augmented reality technologies, applications and limitations. In: International journal of virtual reality, vol. 9(2), pp. 1 (2010)
26. Zhou, F., Duh, H. B. L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. pp. 193-202. IEEE Computer Society (2008)

Antecedents for Cyberloafing – A Literature Review

Katinka Weissenfeld[1], Olga Abramova[2] and Hanna Krasnova[3]

[1] University of Applied Science Bern, Economics, Bern, Switzerland
Katinka.weissenfeld@bfh.ch
[2] Technische Universität Darmstadt, Software Business & Information Management, Darmstadt, Germany
abramova@is.tu-darmstadt.de
[3] University of Potsdam, Business Informatics, Potsdam, Germany
krasnova@uni-potsdam.de

Abstract. The private use of the Internet via desktop and smartphones during working time, also known as cyberloafing, has become a common practice at many workplaces. While critical voices expect performance losses through such behavior, their opponents perceive of the interruptions created by cyberloafing as an opportunity to recover and continue working with increased productivity afterwards. Given the growing body of research on Internet-related employees' opportunism, this paper presents a systematic literature review of 69 studies to identify the factors behind cyberloafing. The classification includes personality traits as well as antecedents related to the job, organization and personal life. The paper concludes with a clear picture of the kind of circumstances which tend to increase cyberloafing and which factors statistically do not seem to have any impact on the abuse of Internet during working time.

Keywords: Cyberloafing, Cyberslacing, Not-work-related computing, Internet addiction.

1 Introduction

Today, most organizations are connected to the Internet to support their routines: to complete electronic payments, to communicate with customers by providing online support, to research new product ideas, to craft and monitor their own brand on Social Media or to collaborate on projects with people all over the world. Despite many ways to boost business efficiency, in many cases the Internet at workplaces is also used by employees for private matters [1-2]. This unproductive use of the Internet during working time is often referred to as cyberloafing, cyberslacking or non-work-related computing. These terms are used as synonyms to signify the abuse of the Internet during working time [3]. Almost 90% of employees misuse the company's Internet access to send and receive private e-mails or visiting news websites [4]. Also shopping (65%), visiting sport websites (60%) and booking vacations (50%) are reasons for employees to go online during work [4]. Furthermore, browsing through Social Media is quite

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

popular and justified by taking mental breaks (34%), connecting with friends and family (27%) and supporting professional connections (24%) [5].

People spend more than one hour per day during eight hours of working time on the Internet [6-7]. As a result, employees continuously interrupt their work which has been found to be more disruptive than external factors [8]. Moreover, for companies the private use of the Internet by employees during working time results in higher expenses [9]. The costs of cyberloafing are at an estimated 85 billion dollars per year for all US companies [7].

Since the boundary between private and business life is continuously more blurry for many people [10], the issue of cyberloafing gains more and more relevance. Employees often require more flexibility at their jobs [11]. On the one hand, people with flexible working conditions, meaning their work is not fixed to certain hours or locations, show higher levels of engagement, stronger organizational commitment and higher job satisfaction [12]. On the other hand, this flexibility implies the use of the Internet for private issues at workplaces.

So far, companies' reactions to cyberloafing are ambiguous. While some companies ban private use of the Internet during working time (e.g. FedEx) [13], others have not taken any actions.

Considering the complexity attached to the concept of cyberloafing, the antecedents of this behavior must be better understood. This study aims to summarize the main factors that drive cyberloafing by conducting a systematic literature review. In terms of research contribution, our paper provides an initial attempt to synthesize existing research findings on the phenomenon. For industry leaders, our work offers a holistic view hinting at how to manage employees in the digital age.

2 Review Method

The literature review follows the guidelines from von Brocke et al. [14] and Webster et al. [15]. Studies were searched using the keyword "cyberloafing" on different scientific databases: Google Scholar (1520); JSTOR (33); ScienceDirect (80); ProQuest (123); InfoTrac (23); ACM Digital Library (2); IEEE Xplore (5); Taylor and Francis (35); Emerald (31) and SAGE journals (24). Synonyms like "cyberslacking" and "non-work-related computing" were also considered during the search process, however no additional unique results were yielded.

After reviewing the titles and abstracts, a total of 231 papers were identified and analyzed in more detail. Out of these, only peer-reviewed papers published in English and with a clear focus on the empirical study of cyberloafing were selected. Since browsing Social Media websites is one of the main reasons for taking a mental break at work [5], the publications' timeframe was set to 2003-2017. This starting point corresponds with the emergence of Social Media websites such as LinkedIn and MySpace, founded in 2002 and 2003 respectively. The selected 145 papers then underwent a full-text review. Finally, a total of 69 studies that investigated the antecedents of cyberloafing were selected as a baseline for further analysis. All studies in the final sample were published between 2004 and 2017 and employ quantitative surveys as the main method. Employees from both private and public sectors took part in the studies. Geographically, the majority of samples originate in the US (50%), Asia (33%) and Turkey (15%). This could be explained by the fact that the work-life balance in these regions is very low, as OECD studies show [16]. Methodologically, 26 studies (37%) apply hierarchical regression for data analysis, and 44 studies adopted different empirical methods such as multiple regression analysis. The results of the analysis are summarized in the following chapter.

3 Results

3.1 Theoretical foundations

Nearly one half of the studies in our sample have a strong theoretical background. In particular, theory of planned behavior, general deterrence theory and social learning theory often serve as conceptual foundation to discover drivers of cyberloafing (Table 1). In some cases, more than one theory was applied. In 26% of the studies the research framework is not rooted in a particular theory and the hypotheses are mainly formed on the basis of past research.

Table 1. Overview of approached theories

N=69	*PBT*	*GDT*	*SLT*	*RCT*	*ET*	*TAT*	*SET*	*BT*	*RT*	*Other*	*No theory*
Number of papers	15 (22%)	12 (17%)	7 (10%)	6 (9%)	5 (7%)	5 (7%)	3 (4%)	3 (4%)	3 (4%)	14 (20%)	18 (26%)

Note: PBT = Planned behavior theory; GDT = General deterrence theory; SLT = Social learning theory; RCT = Rational choice theory; ET = Equity theory; TAT = Trait activation theory; SET = Social exchange theory; BT = Bonder theory; RT = Role theory

The theory of planned behavior states that subjective norms, attitude towards behavior, and perceived behavioral control lead to a deviancy like cyberloafing [17]. In contrast, social learning theory assumes that learning is a cognitive process which happens through the observation of existing norms in groups as well as through rewarding actions or punishing their consequences [18]. In the context of cyberloafing, this means that existing company norms and policies influence employees' behavior. In line with this, general deterrence theory claims individuals can be detained from undesirable acts by using countermeasures like strong disincentives and sanctions [19].

Among others, rational choice theory [20], equity theory [21], trait activation theory [22], social exchange theory [23], border theory [24-25] and role theory [26] were adopted to explain causes of cyberloafing.

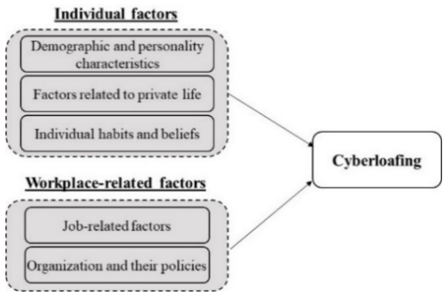


Figure 1. Summary of the antecedents of cyberloafing

Our systematic literature review uncovers several antecedents that underlie individual cyberloafing, as tested and shown by prior research. Considering 83 different original factors, we categorize dominant antecedents into two groups: individual factors and workplace-related factors.

3.2 Individual factors

With regard to individual factors of cyberloafing we distinguish between (1) the demographic and personality characteristics of employees, (2) factors related to their private life, as well as (3) their beliefs and habits with regard to cyberloafing.

Demographic and personality characteristics

Using Internet access for personal purposes while pretending to do legitimate work is linked to a number of different demographic characteristics and personality traits, as presented in Table 2.

Table 2. Demographic and personality characteristics as antecedents of cyberloafing

Demographic and personality characteristics	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Gender (0=male; 1=female)		[4], [28-37]	[6], [38-50]
Age	[28]	[4], [31], [33-34], [36], [41], [43], [46], [51]	[6], [29], [37-38], [42], [44-45], [50], [52-53]
Relationship status (single=1, other=0)	[30], [38]	[54]	
Education	[38-39], [53]	[6], [36], [42], [46], [54], [55]	
Internet skills and computer experiences	[6], [39]	[4], [36], [43-44]	[41]

Demographic and personality characteristics	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Big five personality traits			
- Openness		[38], [45], [56]	
- Conscientiousness		[51], [56], [57]	[38], [45-46], [52], [58], [59]
- Extraversion	[32], [38], [45], [47], [61]	[56]	
- Agreeableness		[38], [45], [51], [56], [57]	[59]
- Neuroticism	[38], [57]	[47], [57]	
Emotional intelligence and honesty		[51], [62]	[36], [45-46], [56], [63], [66]
Self-regulation		[58]	[35], [51], [62]

Previous studies deliver ample evidence that younger [6], [29], [37-38], [42], [44-45], [50], [52-53] and more extroverted people [32], [38], [45], [47], [61], males [6], [38-50], and experienced Internet users [6], [39] tend to cyberloaf more during their working time. Further, a number of studies show that singles are apt to completing job-unrelated tasks at their workplaces [30], [38]. A potential reason for singles to cyberloaf is the search for potential spouses on social networking sites [65]. Additionally, some studies show that high education level [38-39], [53] can be positively related to cyberloafing.

In contrast, high emotional intelligence [36], [45-46], [56], [63], [66], honesty [56], [63], [66], self-regulation [35], [51], [62], conscientiousness [38], [45-46], [52], [58], [59] and agreeableness [59] are personal characteristics which are negatively related to cyberloafing and thus rather contribute to compliant behavior at the workplace.

Factors related to private life

A number of studies manifest that factors related to employees' private life can be linked to cyberloafing (Table 3). For instance, a number of studies show people with many private obligations are apt to engaging in job-unrelated tasks at their workplaces [66-68]. A potential reason might be that people with a high level of private demand tend to use the Internet during working time to organize private matters [65]. Furthermore, previous research ties interruptions during sleep and exhaustion to cyberloafing behavior [47], [69]. One explanation is that sleep interruptions during the night could reduce the intrinsic motivation to work, which in turn leads to more cyberloafing [69].

Table 3. Factors related to private life as antecedents of cyberloafing

Factors related to private life	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Bed time and exhaustion	[47], [69]		
Private demands hours	[66-68]	[70-71]	

Individual habits and beliefs

Interestingly, some employees show a positive attitude towards cyberslacking and perceive of Internet breaks to be useful [4], [61], [72-73] and appropriate [27], [41], [65-66], [70], [72], [74]. Moreover, the power of habituation effect is revealed: those who have integrated cyberloafing into their working routine [27], [33], [35-36], [40], [41-43], [49], [51], [72], [75] practice it more often. In line with social learning theory [18] and theory of planned behavior [17], this witnesses the importance of preventing measures at the initial stage to be able to avert occasional undesirable actions from establishing. The results for individual habits and beliefs as antecedents for cyberloafing are presented in Table 4.

Table 4. Individual habits and beliefs as antecedents of cyberloafing

Individual habits and beliefs	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Personal habits of cyberloafing	[27], [33], [35-36], [40], [41-43], [49], [51], [72], [75]	[4], [29], [31]	
Self-efficacy		[33], [54], [58], [62]	
Normative beliefs and subjective norms	[27], [41], [65-66], [70], [72], [74]	[4], [61]	
Perceived usefulness of cyberloafing	[4], [61], [72-73]		
Procrastination	[76], [65-66]	[57-58]	

3.3 Workplace-related factors

Workplace-related factors of cyberloafing combine reasons attributed to (1) the nature of the job itself, including characteristics the employee doing it, as well as (2) the factors related to the organization and its policies.

Job-related factors

As a subcategory, job-related factors rooted in the nature of the position itself and the function of an employee designated for it are identified from previous work (Table 5).

Table 5. Job-related factors as antecedents of cyberloafing

Job-related factors	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Tenure		[33], [43-44], [51], [77]	
Organizational position	[30], [38], [42], [46], [54-55], [66]	[36], [51], [67]	
Salary/income	[34], [42]	[6], [28], [51], [77]	
Stress and number of working hours	[43], [68], [78-79]	[32], [42], [47], [55], [64]	
Proximity of supervisor	[33], [55]	[38], [67]	[37]
Boredom	[6], [51], [64], [71]		
Meaningfulness of work		[45]	[33], [38], [42], [66-67], [77]
Creativity of work	[6], [66], [80]		
Internet work utility	[6], [42]		
Job satisfaction	[27], [33], [81]	[6], [32], [36], [42], [47], [55], [80]	[41]
Withdrawal behaviors	[80], [82-83]	[30]	

We find that employees in high positions [30], [38], [42], [46], [54-55], [66], with higher income [34], [42] and higher levels of stress [43], [68], [78-79] tend to cyberloaf on the Internet more. In light of cyberloafing as an opportunity to refresh oneself during work, there may be higher need for mental breaks through cyberloafing in jobs that require creativity [6], [66], [80]. Furthermore, studies show that boredom at work [6], [51], [64], [71], the ability to use the Internet to improve the job [6], [42] and the ability to hide cyberloafing at work [83] can provoke cyberloafing. Consequently, those who perceive of their work to be meaningful will engage in cyberloafing significantly less [33], [38], [42], [66-67], [77].

Yet, the links between the level of job satisfaction and cyberloafing behavior remain unclear. Some studies suggest that people tend to cyberloaf [27], [33], [81], if they are satisfied with their jobs. This could be explained also as people which can use the Internet for private matters during working hours are more satisfied with their work. Others disagree this positive correlation and believe employees only substitute their disliked labor with other activities including surfing the Internet [41].

Organizations and their policies

Another cluster of factors contains organizational features, norms and policies (Table 6).

Table 6. Organizational factors as antecedents of cyberloafing

Organizational factors	*Significant positive relationship to cyberloafing*	*No significant relationship to cyberloafing*	*Significant negative relationship to cyberloafing*
Organizational size		[77]	[29], [41], [55]
Norms	[4], [33], [63], [74], [80], [82-83]		
Monitoring and external control		[4], [37], [41], [77]	[29], [31], [36], [44], [52], [73], [84]
Internet usage policies		[41], [55]	[36], [38], [42], [45], [52], [66], [85]
Sanctions		[44], [73]	[27], [37], [43], [63], [85]
Internet access	[38]	[67]	
Organizational justice		[31], [35], [42], [46], [86-87]	[50], [59], [88-89]
Performance appraisal and career advancement		[54]	[90]

It is evident from the research that open access to social network sites [38] and norms which allow Internet use for all purposes [4], [33], [63], [74], [80], [82-83] are significantly associated with cyberloafing.

Furthermore, the bigger the organization [29], [41], [55] and the more thoroughly the monitoring of employees [29], [31], [36], [44], [52], [73], [84], the less likely the Internet is abused during working time. In addition, Internet usage policies [36], [38],

[42], [45], [52], [66], [85] and potential sanctions [27], [37], [43], [63], [85] prevent staff to deviate from actual tasks. These findings are in line with general deterrence theory [19] claiming that awareness of sanctions may decrease a punishable behavior. Furthermore, employees are likely to reduce their opportunistic behavior if they observe peers being penalized [27]. Finally, objective performance ratings, wider perspectives of career opportunities [90] and high levels of organizational justice [50], [59], [88-89] are shown to restrain cyberloafing.

4 Discussion

The aim of this systematic literature review was to investigate the current state of research on the factors behind cyberloafing. A total of 69 papers were analyzed to explore factors associated with this behavior. Our review suggests that theory of planned behavior and general deterrence theory are the most frequently applied theoretical concepts for studying the phenomenon. While theory of planned behavior is mainly focused on determinants from employees' side, like subjective norms, individual attitude, and perceived behavioral control leading to cyberloafing, general deterrence theory is taking the employer perspective with studies centered around preventing measures, policies and sanctions as the main restraining mechanisms. To provide a better overview of factors associated with cyberloafing, two groups of determinants were proposed: individual and workplace-related determinants.

The cluster of individual factors associates cyberslacking with individual characteristics and skills, including gender, age, personality traits, computer skills, sleep quality, relationship status as well as personal attitudes towards cyberloafing. The analysis suggests that younger [6], [29], [37-38], [42], [44-45], [50], [52-53], extraverted people [32], [38], [45], [47], [60] with good computer skills [6], [39] are more likely to abuse access to the Internet for non-work related matters. The second group contains workplace-related determinants like the employee's position within the company, salary and job satisfaction as well as organizational policies on Internet abuse during work time. Employees in higher positions [30], [38], [42], [46], [66], [54-55], their education [38-39], [53], income [34], [42] and levels of stress [43], [68], [81-82] correlate with a higher rate of cyberloafing. Restriction of Internet access [36], [38], [42], [45], [52], [66], [85] and sanctions [27], [37], [43], [63], [85] for non-compliant behavior are shown to be significant measures against cyberloafing.

The literature review illustrates the complexity and multifaceted nature of the cyberloafing phenomenon. On one hand, the given collection of antecedents delineates a typical employee who is inclined to factors which can be (at least partially) influenced by their employer through the number of working hours or the physical proximity between employee and supervisor. On the other hand, our analysis reveals a number of non-workplace-related factors which companies can control less and only influence to a limited extent. Our results are in line with the social trend of work-life-blending [93], which means that the boundaries between work and life are increasingly softened which has employees mix their professional and private interests. Especially employees who need to spend much time at work [43], [68], [78-79] use cyberloafing as a way to follow up on private matters during working time [29]. Furthermore, our findings identify that young people [6], [29], [37-38], [42], [44-45], [50], [52-53] and employees of small companies [29], [41], [55] tend to cyberloaf more. These results are also conform to work-life-blending trends. Especially young people (the so-called Generation Y) and employees of start-up companies do not draw clear lines between work and private life [94].

Although organizations have an interest in reducing cyberloafing, as the employees' productivity could be influenced in a negative way [95], we have found enough confounding evidence that the breaks taken to cyberloaf are helpful in providing inspiration for creative work [6], [66], [80], increasing job satisfaction [27], [33], [81] and reducing work-related stress [92].

Our results further imply that the antecedents of cyberloafing have many different business specificities, and the strategies of handling cyberslacing depend heavily on the corporate culture and governance of the respective company.

5 Conclusion, limitations and future research

Cyberloafing at the workplace has become common in the digital age. Summarizing extant research, this paper provides a structured review of current literature on cyberloafing with a special focus on its antecedents. However, empirical results remain scattered. This paper addresses this gap by conducting a systematic literature review and providing a comprehensive summary of existing findings on the factors behind cyberloafing. Furthermore, this study extends IS research by presenting a number of practical implications for corporations, their management and employees. We reveal an array of conflictual findings that exist in the literature, which calls for a more thorough exploration into the reasons of these diverging insights.

The current study is subject to a number of limitations. First, the papers included in our sample were found through keyword search, followed by a subsequent exclusion. Second, we acknowledge possible interdependences of factors in our classification which is mainly attributed to the heterogeneity of the studies summarized. For example, age could be a confounding variable in the correlation between a) computer skills and cyberloafing and b) relationships status and cyberloafing. This is because younger generations have better digital literacy and are not likely to have founded own families yet. Therefore, our framework compiles and recaps previous research without delivering a strong causal model. Third, we treated cyberloafing as a unified construct without analyzing its multiple facets like type of activity or time of occurrence. Recognizing these limitations, our research serves as point of departure for future investigations in this area. For instance, future research could take into account the organizational culture and the corresponding work-life-blending, industry type (e.g. construction vs. IT sector), type of tasks (e.g. creative vs. routine jobs) and types of employment (full time, part time, casual, fixed term, apprentices and trainees). Moreover, intercultural comparisons between Western and Eastern countries would be interesting.

References

1. Jandaghi, G, Alvani, S. M., Matin, H. Z., Kozekanan, S., F.: Cyberloafing Management in Organizations. Iranian Journal of Management Studies. 8:3, 335-349 (2015)
2. Hussain, S., Prida, T.: Exploring cyberloafing behavior in South-central Ethiopia: A close look at Madda Walabu University. Journal of Media and Communication Studies. 9:2, 10-16 (2017)
3. Lim, V. K. G.: The IT way of loafing on the job: Cyberloafing, neutralizing and organizational justice. Journal of Organizational Behavior. 23, 675–694 (2002)
4. Blanchard, A., Henle, C.: Correlates of different forms of cyberloafing: The role of norms and external locus of control. Computers in Human Behavior. 24, 1067–1084 (2008)
5. Olmstead, K., Lampe, C., Ellison, N. B.: Social Media and the Workplace. Pew Research Center, 2-14 (2016)
6. Vitak, J., Crouse, J., LaRose, R.: Personal Internet use at work: Understanding cyberslacking. Computers in Human Behavior. 27, 1751–1759 (2011)
7. Fox, A.: Taking Up Slack: Most slackers can be turned into better performers by removing organizational conditions that create or enable loafing behavior. HR Magazine. 55:12, 26-31 (2010)
8. Katidioti, I., Borst, J. P., van Vugt, M. K., Taatgen, N. A.: Interrupt me: External interruptions are less disruptive than self-interruptions. Computers in Human Behavior. 63, 906-915 (2016)
9. Weatherbee, T. G.: Counterproductive use of technology at work: Information & communications technologies and cyberdeviancy. Human Resource Management Review. 20:1, 35-44 (2010)
10. Köffer, S., Anlauf, L., Ortbach, K., Niehaves, B.: The Intensified Blurring of Boundaries Between Work and Private Life through IT Consumerisation. ECIS (2015)
11. Kultalahti, S.: It's so nice to be at work! Adopting different perspectives in understanding Generation Y at work. University of Vaasa, Finnland (2015)
12. Richman, A. Johnson, A., Noble, K.: Business Impacts of Flexibility: An Imperative for Expansion. Corporate Voices for Working Families, Washington (2011)
13. Perkins, A.: Should Employers Restrict Personal Device Use at Work?. Common Places Culture Strategy, https://www.commonplaces.com/blog/should-employers-restrict-personal-device-use-at-work/ (Accessed: 14.08.2018)
14. von Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. Proceedings of the 17th European Conference on Information Systems. 9, 2206–2217 (2009)
15. Webster, J., Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly. 26:2, xiii–xxiii (2002)
16. OECD, Better Life Index 2017, http://www.oecdbetterlifeindex.org/topics/work-life-balance/ (Accessed: 14.08.2018)
17. Ajzen, I.: From intentions to actions: A theory of planned behavior. In Kuhl, J. Beckmann, J. (eds.), Action-control: From cognition to behavior. pp. 11-39. Springer, Heidelberg (1985)
18. Bandura, A.: Social Learning Theory. General Learning Press, New York (1971)
19. Schuessler, J.: General deterrence theory: Assessing information systems security effectiveness in large versus small businesses. Ph.D. dissertation, University of North Texas, United States, Texas (2009)
20. Scott, J.: Rational Choice Theory. In Browning, G., Halcli., A., Webster, F. (eds.), Understanding Contemporary Society: Theories of The Present, Sage Publications (2000)

21. Adams, J. S.: Inequity in social exchange. In Berkowitz, L. (eds), Advances in experimental social psychology. Academic Press, vol. 2, pp. 267-299. New York (1965)
22. Tett, R. P., Burnett, D. D.: A personality trait-based interactionist model of job performance. Journal of Applied Psychology. 88, 500–517 (2003)
23. Blau, P. M.: Exchange and power in social life. Wiley, pp. 88-115. New York (1964)
24. Ashforth, B. E., Kreiner, G. E., Fugate, M.: All in a day's work: boundaries and micro role transitions. Academy of Management Review. 25:3, 472-491 (2000)
25. Clark, S. C.: Communicating across the work/home border. Community, Work & Family. 5, 23–48 (2002)
26. Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., Rosenthal, R. A.: Organizational Stress: Studies in Role Conflict and Role Ambiguity, John Wiley & Sons, New York (1964)
27. Woon, I. M. Y., Pee, L. G.: Behavioral Factors Affecting Internet Abuse in the Workplace – An Empirical Investigation. Proceedings of the Third Annual Workshop on HCI Research in MIS. 80-84, Washington DC (2004)
28. Akman I., Mishra, A.: Gender, age and income differences in internet usage among employees in organizations. Computers in Human Behavior. 26, 482–490 (2010)
29. Banerjee, S., Thakur, S.: A Critical Study of Factors Promoting Cyberloafing in Organizations. ICTCS '16 Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. 139, 1-6 (2016)
30. Çınar, O., Karcioglu, F.: The relationship between cyber loafing and organizational citizenship behavior: A survey study in Erzurum/Turkey. Procedia - Social and Behavioral Sciences. 207, 444-453 (2015)
31. Jamaluddin, H., Ahmad, Z., Alias, M., Simun, M., Personal Internet use: The use of personal mobile devices at the workplace. Procedia - Social and Behavioral Sciences. 172, 495-502 (2015)
32. Krishnan, S., Lim, V., Teo, T.: How Does Personality Matter? Investigating the Impact of Big Five Personality Traits on Cyberloafing. ICIS 2010 Proceedings. Paper 6 (2010)
33. Liberman, B., Seidman, G., McKenna, K. Y. A., Buffardi, L. E.: Employee job attitudes and organizational characteristics as predictors of cyberloafing. Computers in Human Behavior. 27, 2192–2199 (2011)
34. Rajah, R., Lim, V. K. G.: Cyberloafing, Neutralization, And Organizational Citizenship Behavior. PACIS 2011 Proceedings. 152, 1-15 (2011)
35. Restubog, S. L., Garcis, P. R., Toledano, L. S., Amarnani, R. K., Tolendino, L. R., Tang, R. L.: Yielding to (cyber)-temptation: Exploring the buffering role of self-control in the relationship between organizational justice and cyberloafing behavior in the workplace. Journal of Research in Personality. 45, 247–251 (2011)
36. Wang, J. Tian, J., Shen, Z.: The effects and moderators of cyber-loafing controls: an empirical study of Chinese public servants. Information Technology and Management. 14, 269–282 (2013)
37. Zoghbi-Manrique-de-Lara, P., Olivares-Mesa, A.: Bringing cyber loafers back on the right track. Industrial Management & Data Systems. 110:7, 1038-1053 (2010)
38. Andreassen, C., Torsheim, T., Pallesen, S.: Predictors of Use of Social Network Sites at Work - A Specific Type of Cyberloafing. Journal of Computer-Mediated Communication. 19, 906–921 (2014)
39. Baturaya, M., Tokar, S.: An investigation of the impact of demographics on cyberloafing from an educational setting angle. Computers in Human Behavior.50, 358–366 (2015)
40. Carlson, J., Zivnuska, S., Harris, R., Harris, K.: Social Media Use in the Workplace: A Study of Dual Effects. Journal of Organizational and End User Computing. 28:1, 15-31 (2016)

41. Galletta, D., Polak, P.: An Empirical Investigation of Antecedents of Internet Abuse in the Workplace. SIGHCI 2003 Proceedings. 14, 47-51 (2003)
42. Garrett, K., Danziger, J. N.: Disaffection or expected outcomes: Understanding personal Internet use during work. Journal of Computer-Mediated Communication. 13, 937-958 (2008)
43. Henle, C. A., Blanchard, A. L.: The Interaction of Work Stressors and Organizational Sanctions on Cyberloafing. Journal of Managerial Issues. 20:3, 383-400 (2008)
44. Henle, C. A, Kohut, G, Booth, R.: Designing electronic use policies to enhance employee perceptions of fairness and to reduce cyberloafing: An empirical test of justice theory. Computers in Human Behavior. 25, 902–910 (2009)
45. Jia, H., Jia, R., Karau, S.: Cyberloafing and Personality: The Impact of the Big Five Traits and Workplace Situational Factors. Journal of Leadership & Organizational Studies. 20:3, 358-365 (2013)
46. Kim, K., Triana, M., Chung, K., Oh, N.: When do Employees Cyberloaf? An Interactionist Perspective Examining Personality, Justice and Empowerment. Human Resource Management. 55: 6, 1041–1058 (2016)
47. Krishnan, S., Lim, V.: Moderating Effects of Extroversion and Neuroticism on Sleep Deprivation and Cyberloafing. PACIS 2010 Proceedings. 100, 1020-1031 (2010)
48. Lim, V. K. G., Chen, D. J. Q.: Cyberloafing at the workplace: gain or drain on work?. Behaviour & Information Technology. 31:4, 343-353 (2012)
49. Yılmaz, F. G., Yılmaz, R., Öztürk, H. T., Sezer, B., Karademir, T.: Cyberloafing as a barrier to the successful integration of information and communication technologies into teaching and learning environments. Computers in Human Behavior. 45, 290–298 (2015)
50. Zoghbi-Manrique-de-Lara, P.: Relationship between Organizational Justice and Cyberloafing in the Workplace: Has "Anomia" a Say in the Matter?. CyberPsychology & Behavior. 10:3, 464-470 (2007)
51. Mercado, B. K., Giordano, C., Dilchert, S.: A meta-analytic investigation of cyberloafing. Career Development International. 22:5, 546-564 (2017)
52. Hassan, H. M., Reza, D. M., Farkhad, M.: An Experimental Study of Influential Elements on Cyberlaofing from General Deterrence Theory Perspective. International Business Research. 8:3, 91-98 (2015)
53. Phillips, J. G., Reddie, L.: Decisional style and self-reported Email use in the workplace. Computers in Human Behavior. 23, 2414–2428 (2007)
54. Hanna, B., Kee, K., Robertson, B.: Positive Impacts of Social Media at Work: Job Satisfaction, Job Calling, and Facebook Use among Co-Workers. SHS Web of Conferences i-COME'16. 33, 1-7 (2017)
55. Wyatt, K., Phillips, J. G.: Internet Use and Misuse in the Workplace. Proceedings of OZCHI, Canberra, Australia (2005)
56. Seymour, L., Nadasen, K.: Web access for IT staff: a developing world perspective on web abuse. The Electronic Library. 25:5, 543-557 (2007)
57. Wagner, D. T., Barnes, C. M., Lim, V. K. G., Ferris, D. L.: Lost Sleep and Cyberloafing: Evidence From the Laboratory and a Daylight Saving Time Quasi-Experiment. Journal of Applied Psychology. Advance online publication. 1-9 (2012)
58. Yan, J., Yang, J.: Trait procrastination and compulsive Internet use as predictors of cyberloafing. 11th International Conference on Service Systems and Service Management (2014)
59. Cheng, L., Li, W., Zhai, Q., Smyth, R.: Understanding personal use of the Internet at work: An integrated model of neutralization techniques and general deterrence theory. Computers in Human Behavior. 38, 220–228 (2014)

60. Sheikh, A., Atashgah, M. S., Adibzadegan, M.: The antecedents of cyberloafing: A case study in an Iranian copper industry. Computers in Human Behavior. 51, 172–179 (2015)
61. RuningSawitri, H. S.: Role of Internet Experience in Moderating Influence of Work Stressor on Cyberloafing. Procedia - Social and Behavioral Sciences. 57, 320 – 324 (2012)
62. Wana, H. C., Downey, L. A., Stough, C.: Understanding non-work presenteeism: Relationships between emotional intelligence, boredom, procrastination and job stress. Personality and Individual Differences. 65, 86–90 (2014)
63. Koay, K. Y., Soh, P. C., Chew, K. W.: Do employees' private demands lead to cyberloafing? The mediating role of job stress. Management Research Review. 40:9, 1025-1038 (2017)
64. König, C. J., Caner de la Guardia, M. E.: Exploring the positive side of personal internet use at work: Does it help in managing the border between work and nonwork?. Computers in Human Behavior. 30, 355–360 (2014)
65. Fichtner, J. R., Strader, T. J.: Non-Work-Related Computing and Job Characteristics: Literature Review and Future Research Directions. Journal of Psychological Issues in Organizational Culture. 4:4, 65-79 (2014)
66. Yasar, S., Yurdugül, H.: The Investigation Of Relation Between Cyberloafing Activities And Cyberloafing Behaviors In Higher Education. Procedia - Social and Behavioral Sciences. 83, 600–604 (2013)
67. Huma, Z., Hussain, S., Thurasamy, R., Malik, M. I.: Determinants of cyberloafing: a comparative study of a public and private sector organization. Internet Research. 27:1, 97-117 (2017)
68. TU, X., Chang, Y.: An Empirical Study on the Determinants of Cyberloafing: Data Analysis Based on Individual Factors. International Conference on E-Product E-Service and E-Entertainment. 1-4 (2010)
69. Jafarkarimi, H., Saadatdoost, R., Sim, A. T., Hee J. M.: Behavioral intention in social networking sites ethical dilemmas: An extended model based on Theory of Planned Behavior. Computers in Human Behavior. 62, 545-561 (2016)
70. Andreassen, C., Torsheim, T., Pallesen, S.: Use of online social network sites for personal purposes at work: does it impair self-reported performance?. Comprehensive Psychology. 3, 1-11 (2014)
71. Koay, K. Y., Soh, P. C., Chew, K. W.: Antecedents and consequences of cyberloafing: Evidence from the Malaysian ICT industry. First Monday. 22:3 (March 6th 2017)
72. Lee, O. D., Lim, K. H., Wong, W. M.: Why Employees Do Non-Work-Related Computing: An Exploratory Investigation through Multiple Theoretical Perspectives. Proceedings of the 38th Hawaii International Conference on System Sciences. (2005)
73. Askew, K., Buckner, J., Taing, M., Ilie, A., Bauer, J., Coovert, M.: Explaining cyberloafing: The role of the theory of planned behavior. Computers in Human Behavior. 36, 510–519 (2014)
74. Prasad, S., Lim, V. K. G., Chen, D. J. Q.: Self-Regulation, Individual Characteristics and Cyberloafing. PACIS 2010 Proceedings. 159,1641-1648 (2010)
75. O'Neill, T. A., Hambley, L. A., Bercovich, A.: Prediction of cyberslacking when employees are working away from the office. Computers in Human Behavior. 34, 291–298 (2014)
76. Gökçearslan, S., Mumcu, F. K., Haslaman, T., Cevik, Y. D.: Modelling smartphone addiction: The role of smartphone usage, selfregulation, general self-efficacy and cyberloafing in university students. Computers in Human Behavior. 63, 639-649 (2016)
77. Yang, J., Diefendorff, J. M.: The Relations of Daily Counterproductive Workplace Behavior with Emotions, Situational Antecedents, and Personality Moderators: A Diary Study in Hong Kong. Personnel Psychology. 62, 259–295 (2009)

78. O'Neill, A. T., Hambley, L. A., Chatellier, G. S.: Cyberslacking, engagement, and personality in distributed work environments. Computers in Human Behavior. 40, 152–160 (2014)
79. Eddy, E. R., D'Abate, C. P., Thurston Jr, P. W.: Explaining engagement in personal activities on company time. Personnel Review. 39:5, 639-654 (2010)
80. Bock, G., Park, S., Zhang, Y.: Why Employees Do Non-Work-Related Computing in the Workplace. Journal of Computer Information Systems. 50:3, 150-163 (2010)
81. Lee, A., Bruckman, A.: Judging you by the company you keep: Dating on social networking sites. Assoc Computing Machinery, New York (2007)
82. Lowry, P. B., Posey, C., Bennett, R. J., Roberts, T. L.: Leveraging fairness and reactance theories to deter reactive computer abuse following enhanced organisational information security policies: an empirical study of the influence of counterfactual reasoning and organisational trust. Information Systems Journal. 25, 193–230 (2015)
83. Aghaz A., Sheikh, A.: Cyberloafing and job burnout: An investigation in the knowledge-intensive sector. Computers in Human Behavior. 62, 51-60 (2016)
84. Khansa, L., Kuem, J., Siponen, M., Kim, S.: To Cyberloaf or Not to Cyberloaf: The Impact of the Announcement of Formal Organizational Controls. Journal of Management Information Systems. 34:1, 141-176 (2017)
85. Ugrin, J. C., Pearson, M.: The effects of sanctions and stigmas on cyberloafing. Computers in Human Behavior. 29, 812–820 (2013)
86. Li, H., Sarathy, R., Zheng, J., Luo, X.: Exploring the effects of organizational justice, personal ethics and sanction on internet use policy compliance. Information Systems Journal. 24, 479–502 (2014)
87. Lim, V. K. G.: The Moderating Effect of Neutralization Technique on Organizational Justice and Cyberloafing. PACIS 2005 Proceedings. 18, 207-219 (2005)
88. Son, J.-Y., Park, J.: Procedural justice to enhance compliance with non-work-related computing (NWRC) rules: Its determinants and interaction with privacy concerns. International Journal of Information Management. 36, 309–321 (2016)
89. Willison, R., Warkentin, M., Johnston, A. C.: Examining employee computer abuse intentions: insights from justice, deterrence and neutralization perspectives. Information Systems Journal. 28:2, 266-293 (2016)
90. Al-Shuaibi, A., Shamsudin, F., Subramaniam, C.: Do Human Resource Management Practices Matter in Reducing Cyberloafing at Work: Evidence from Jordan. Journal of WEI Business and Economics. 2:2, 1-11 (2013)
91. Marlin Company and American Institute of Stress: Attitudes in the American Workforce VII (2001)
92. Lavoie, J. A. A., Pychyl, T. A.: Cyberslacking and the Procrasination Superhighway: A Web-based Survey of Online Procrastiantion, Attitures, and Emotion. Social Science Computer Review.19, 431-444 (2001)
93. Campbell Clark, S., Work / family border theory: A new theory of work / family balance. Human Relations. 53:6, 747-440 (2000)
94. Smith, K. T.: Work-Life Balance Perspectives of Marketing Professionals in Generation Y. Services Marketing Quarterly. 31:4, 434-447 (2010)
95. Duke, E., Montag, C.: Smartphone addiction, daily interruptions and self-reported productivity. Addictive Behaviors Reports. 6, 90-95 (2017)

Internal Crowd Work as a Source of Empowerment - An Empirical Analysis of the Perception of Employees in a Crowdtesting Project -

David Durward[1], Benedikt Simmert[1], Ivo Blohm[2], and Christoph Peters[1,2]

[1] University of Kassel, Information Systems, Kassel, Germany
{david.durward,benedikt.simmert,christoph.peters}@uni-kassel.de
[2] University of St.Gallen, Institute of Information Management, St.Gallen, Switzerland
{ivo.blohm,christoph.peters}@unisg.ch

Abstract. Internal crowd work has emerged as a new form of digital gainful employment that changes the nature of work. However, the possible effects of internal crowd work on the individual level have been largely neglected. In this paper, we therefore present our research in progress which is concerned with the effects of work characteristics in internal crowd work that have impact on the individual's empowerment and satisfaction. Thus, we developed our research model and conducted an online survey amongst 118 internal crowd workers of a Swiss bank who were asked to test new software. Our expected contribution will increase the understanding of internal crowd work and provide important insights for organizations to (re-) design work on internal IT-platforms.

Keywords: Crowdsourcing, Internal Crowd Work, Empowerment, Work Characteristics.

1 Introduction

Companies increasingly use IT-platforms to reach out to employees, speed up collaborative work and support ideation within the organization [1]. As a new form of collaboration, internal crowd work reflects an IT-enabled group activity based on an open call for participation in an enterprise [2]. While external crowd work is performed with individuals from outside the company boundaries, in internal settings, the own employees act as an internal crowd to leverage the collective intelligence as part of their working time [3]. Internal crowd work has seen a substantial uptake in practice and has attracted a first wave of research papers and dedicated studies [2]. There is abundant research that exhibits the potentials of internal crowd work for organizations, such as fast access to internal knowledge [2, 4]. However, the perspective of employees – the individuals who act as internal crowd workers – has been largely neglected. Although few studies addressed the motivations of individuals working in the crowd, there is a gap in understanding experiences and perceptions of internal crowd workers [5]. While prior studies focus on efficiency and performance dimensions, systematic analysis regarding the intended or unintended ethical consequences of these new work structures

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

have been mostly absent [5]. Therefore, it is essential to better understand internal crowd work as a new type of digital work out of an employee's perspective. One well-established construct that is associated with the individual perception of work is empowerment which enhances work satisfaction [6]. Since prior studies shows that employees working in collaborative environments tend to be more satisfied [7], we believe similar effects to be present in internal crowd work. However, while satisfaction in crowd work has so far been examined mainly as a motivational construct [e.g., 8], we thus analyze the crowd workers' perceptions with regard to work itself to reconceptualize the phenomenon [9]. Thus, this survey seeks to fill the outlined research gaps by addressing the following research question: *How does empowerment affect the employees' satisfaction with internal crowd work?*

2 Theoretical Background

As theoretical lens of our paper, we apply the self-determination theory (SDT) that evolved from research on intrinsic as well as extrinsic motivations and expanded to include research on work organizations [10]. In general, SDT suggests that both individuals' performance and their well-being are affected by the type of motivation they have for their job activities [10]. There are several concepts of SDT that have been extended to work characteristics literature as important aspects of individuals' work [11]. In particular, work characteristics provide several core objective features (e.g., task, variety, task significance, feedback, or task complexity) associated with any job and further drive psychological states such as self-determination [12]. In this context, Seibert et al. [13] found work characteristics to be contextual antecedents of empowerment. In general, the psychological construct of empowerment examines individual experiences and intrinsic motivational aspects of the employee [14]. In particular, the subjective and individual interaction of the employees with the given work structures is investigated [15]. Against this background, the organizational structures (i.e., work characteristics) influence the individual interpretations of psychological empowerment.

3 Research Model and Hypothesis Development

Drawing on SDT [10], we unravel the mechanisms of how work characteristics affect satisfaction in internal crowd work. We assume psychological empowerment to be a mediator and thus to act as a mechanism through which the work characteristics influence satisfaction. On the one hand, internal crowd work initiatives provide a range of different tasks to be performed. On the other hand, due to the platforms automated evaluation and review processes, crowd workers receive direct feedback on their performance. Further, internal crowd work reflects an opportunity for employees to collaboratively work on more complex projects. Thus, we focused on *task variety*, *feedback from task*, and *problem solving* as work characteristics that are grounded in the very nature of internal crowd work. We expect employees who perceive high task variety, feedback from task and problem solving due to their participation to feel more empowered and further satisfied. In our case, we developed our research model in the context of an internal crowd work project of a Swiss bank. The bank is currently developing new software and usually integrates their employees in the testing of its internal applications. In sum, figure 1 depicts our research model.

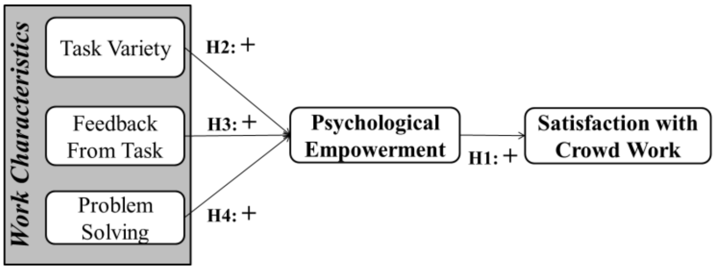


Figure 1. Research model

Previous research [12-15] has shown, that psychological empowerment can be regarded as mediating variable between structural work characteristics and job satisfaction. In this study, we assume similar effects since the voluntary testing of special internal software or applications enhance a feeling of self-determination and thus empower the employees. In turn, the employees might feel more satisfied due to their newly acquired freedom in task performance. Hence, we assume:

H1) *A crowd workers' psychological empowerment positively influences the satisfaction.*

As already mentioned, work characteristics seem to be associated with psychological empowerment [13]. Thus, we first introduce task variety that refers to the degree to which a job requires workers to perform a wide range of tasks [11]. Jobs of high task variety are more likely to empower subordinates since they break the chains of monotony [16]. In our case, the employees choose between various testing tasks (e.g., integration tests, interface tests, or security tests) that increase self-determination in daily work and foster satisfaction. Hence, we expect:

H2) *The positive effect of task variety on satisfaction is mediated by empowerment.*

The second essential characteristic is feedback from task that reflects the degree to which jobs provide direct and clear information about the effectiveness of task performance [17]. In this study, the employees receive direct feedback after the submission of their bug reports via the internal IT-platform. Thus, they are immediately informed about their testing performance and are better able to plan necessary next steps themselves. As a result, they are even more satisfied, which is why we suggest:

H3) *The positive effect of feedback from task on satisfaction is mediated by empowerment.*

Third, problem solving refers to the degree to which a job requires unique ideas or solutions and reflects the more active cognitive processing requirements of a job [18]. Jobs with high problem-solving requirements provide a chance for employees to perform in challenging, novel situations in which they are able to demonstrate their sense of competence [19]. In internal crowd work, the testing requires specific expert knowledge of employees [20]. In our study, the internal testers were asked to find all kinds of bugs within the new software. Since no best solution is specified in advance, the employees independently conduct explorative analysis. This procedure gives meaning to the testing tasks and thereby empower the employees. As a result, solving challenging tests might further increase employee's satisfaction. Thus, we assume:

H4) *The positive effect of problem solving on satisfaction is mediated by empowerment.*

4 Research Method

We conducted an IT-based survey to collect data from internal crowd workers within a Swiss bank. In sum, a total of 118 internal crowd workers provided a completed questionnaire (response rate: 46.8%). We used well-established and valid items for the measurement of work characteristics [11], psychological empowerment [14], and satisfaction [21]. To minimize extreme response and acquiescence biases, we include items with both positive and negative wording [22, 23]. We have already applied exploratory and confirmatory factor analysis to confirm validity and reliability of our measures. The measure of sampling adequacy was 0.78, indicating excellent applicability of exploratory factor analysis. We extracted 5 clearly interpretable factors. Since the values for composite reliabilities and the average variance explained surpassed the value of 0.5, convergent validity could be assumed [24]. As part of further analysis, we assume a mediation effect that describes when and under what conditions an effect occurs. Thus, we will perform ordinary least squares (OLS) regressions with a nonparametric bootstrapping approach to compute bias-corrected confidence intervals for testing our hypothesis [25, 26]. This allows us to investigate our research model while ruling out alternative models [27].

5 Expected Contributions

To our knowledge, this study is the first to investigate the nature of internal crowd work from an individual's perspective and therefore provides two main contributions. First, we gain information for a better understanding of internal crowd work [5, 28]. According to Rivard [9], we conceptually clarify the construct of psychological empowerment by examining its antecedents and effects on satisfaction, based on the individual perceptions of employees. Second, in line with prior research on SDT [10, 29], we unravel different work characteristics (i.e., *task variety, feedback from task,* and *problem solving*) as predictors of empowerment in internal crowd work. Thus, we illustrate that organizations can empower and even enhance satisfaction among their workforce by implementing internal crowd work. In sum, we believe that the employees might feel more self-determined and free in daily work due to their participation.

6 Acknowledgements

This paper was written as part of the projects "*Die interne Crowd*", funded by the Hans Böckler Stiftung (HBS) and "*Empowerment in der digitalen Arbeitswelt (EdA)*" (funding number: 02L15A051), funded by the German Federal Ministry of Education and Research (BMBF) and the European Social Fund (ESF).

References

1. Dery, K., Sebastian, I.M., van der Meulen, N.: The Digital Workplace is Key to Digital Innovation. MIS Quarterly Executive 16, (2017)
2. Zuchowski, O., Posegga, O., Schlagwein, D., Fischbach, K.: Internal Crowdsourcing: Conceptual Framework, Structured Review and Research Agenda. Journal of Information Technology (2016)
3. Durward, D., Blohm, I., Leimeister, J.M.: Crowd work. Business & Information Systems Engineering 58, 281-286 (2016)
4. Malhotra, A., Majchrzak, A., Kesebi, L., Looram, S.: Developing innovative solutions through internal crowdsourcing. MIT Sloan management review 58, 73 (2017)
5. Deng, X., Joshi, K., Galliers, R.D.: The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. Management Information Systems Quarterly 40, 279-302 (2016)
6. Liden, R.C., Wayne, S.J., Sparrowe, R.T.: An examination of the mediating role of psychological empowerment on the relations between the job, interpersonal relationships, and work outcomes. Journal of applied psychology 85, 407 (2000)
7. Scott, S.G., Bruce, R.A.: Determinants of innovative behavior: A path model of individual innovation in the workplace. Academy of Management journal 37, 580-607 (1994)
8. Afuah, A., Tucci, C.L.: Crowdsourcing as a solution to distant search. Academy of Management Review 37, 335-375 (2012)
9. Rivard, S.: Editor's comments: the ions of theory construction. Mis Quarterly 38, iii-xiv (2014)
10. Deci, E.L., Olafsen, A.H., Ryan, R.M.: Self-determination theory in work organizations: The state of a science. Annual Review of Organizational Psychology and Organizational Behavior 4, 19-43 (2017)
11. Morgeson, F.P., Humphrey, S.E.: The work design questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. Journal of Applied Psychology 91, 1321 (2006)
12. Maynard, M.T., Gilson, L.L., Mathieu, J.E.: Empowerment—fad or fab? A multilevel review of the past two decades of research. Journal of Management 38, 1231-1281 (2012)
13. Seibert, S.E., Wang, G., Courtright, S.H.: Antecedents and consequences of psychological and team empowerment in organizations: a meta-analytic review. American Psychological Association (2011)
14. Spreitzer, G.M.: Psychological empowerment in the workplace: Dimensions, measurement, and validation. Academy of Management journal 38, 1442-1465 (1995)
15. Spreitzer, G.M.: Taking stock: A review of more than twenty years of research on empowerment at work. Handbook of organizational behavior 54-72 (2008)
16. Conger, J.A., Kanungo, R.N.: The empowerment process: Integrating theory and practice. Academy of management review 13, 471-482 (1988)
17. Hackman, J.R., Oldham, G.R.: Work redesign. (1980)
18. Jackson, P.R., Wall, T.D., Martin, R., Davids, K.: New measures of job control, cognitive demand, and production responsibility. Journal of applied Psychology 78, 753 (1993)
19. Deci, E.L., Ryan, R.M.: The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry 11, 227-268 (2000)
20. Leicht, N., Rhyn, M., Hansbauer, G.: Can Laymen Outperform Experts? The Effects of User Expertise and Task Design in Crowdsourced Software Testing. In: Conference Can Laymen Outperform Experts? The Effects of User Expertise and Task Design in Crowdsourced Software Testing. (Year)

21. van Dick, R., Wagner, U.: Social identification among school teachers: Dimensions, foci, and correlates. European Journal of Work and Organizational Psychology 11, 129-149 (2002)
22. Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P.: Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 88, 879 (2003)
23. Sauro, J., Lewis, J.R.: When designing usability questionnaires, does it hurt to be positive? In: Conference When designing usability questionnaires, does it hurt to be positive?, pp. 2215-2224. ACM, (Year)
24. Bagozzi, R.P., Yi, Y.: On the evaluation of structural equation models. Journal of the Academy of Marketing Science 16, 74-94 (1988)
25. Hayes, A.F.: Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. Guilford Press (2013)
26. Preacher, K.J., Rucker, D.D., Hayes, A.F.: Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. Multivariate Behavioral Research 42, 185-227 (2007)
27. Muller, D., Judd, C.M., Yzerbyt, V.Y.: When moderation is mediated and mediation is moderated. Journal of Personality and Social Psychology 89, 852 (2005)
28. Kittur, A., Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J.: The future of crowd work. In: Conference The future of crowd work, pp. 1301-1318. ACM, (Year)
29. Gagné, M., Senecal, C.B., Koestner, R.: Proximal job characteristics, feelings of empowerment, and intrinsic motivation: A multidimensional model 1. Journal of Applied Social Psychology 27, 1222-1240 (1997)

Track 14:

Geschäftsmodelle und digitales Unternehmertum

Track Chairs:

Prof. Dr. Daniel Veit
University of Augsburg

Prof. Dr. Alexander Benlian
Darmstadt University of Technology

Dr. Benedikt Berger
Ludwig-Maximilians Universität München

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Dividing the ICO Jungle: Extracting and Evaluating Design Archetypes

Nina Bachmann[1], Benedict Drasch[2,3], Michael Miksch[1], André Schweizer[2,3]

[1] FIM Research Center, University of Augsburg, Augsburg, Germany
{nina.bachmann,michael.miksch}@tum.de
[2] Project Group Business & Information Systems Engineering of Fraunhofer FIT, Germany
{benedict.drasch, andre.schweizer}@fit.fraunhofer.de
[3] FIM Research Center, University of Bayreuth, Bayreuth, Germany
{benedict.drasch, andre.schweizer}@fim-rc.de

Abstract. The sale of blockchain-based digital tokens as a novel funding mechanism, referred to as initial coin offerings (ICO), has grown exponentially, resulting in $12bn raised globally during the first half of 2018. Due to the novelty of the phenomenon, the concept is not yet entirely understood. Existing research provides first insights into ICO endeavors and design only. To date, comprehensive and in-depth analyses of ICO design archetypes to better understand prevailing ICO characteristics are missing. We bridge this gap by enriching an existing ICO taxonomy and applying a cluster analysis to identify predominant ICO archetypes. As a result, we identify five ICO design archetypes: the average ICO, the liberal ICO, the visionary ICO, the compliant ICO, and the native ICO. We thereby contribute to a comprehensive and in-depth understanding of the ICO phenomenon and its implications. Further, we offer practitioners tangible design suggestions for future ICOs.

Keywords: Blockchain, initial coin offering, ICO, cluster analysis, design archetypes

1 Introduction

Emerging digital technologies challenge existing business structures and invoke innovation [1, 2]. As one example, blockchain forces organizations to rethink and innovate their business models. Thus, while the technology's potential is not yet entirely assessed and understood, we observe increasing interest in its vast use cases from both practitioners and academics [3, 4]. In the past years, a use case in the financial service industry is attracting high attention: sales of blockchain-based digital tokens as a novel funding mechanism, referred to as initial coin offerings (ICOs) [5-9]. Despite regulatory uncertainty [10-12], ICO fundraising has grown exponentially throughout 2017 (343 ICOs) and 2018 (394 ICOs in six months) [13]. Indeed, for the first half of 2018, the Wall Street Journal reports $12bn raised in the global ICO market [14].

Due to the novelty of the phenomenon, the concept of ICOs is not yet entirely understood [5], and a number of questions - especially related to regulation - need to be

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

answered in practice and academia. With regard to the ICO's inherent idea of providing open, global, and decentralized access to funding, regulation becomes very difficult [12]. Regulators and many governmental institutions have just started to take action in the so far mostly unregulated ICO market [7]. The regulation approaches, however, are neither homogeneous, nor follow an integrated global strategy. Thus, the actions range from banning ICOs to taking no action or focusing on specific ICOs only [15]. One major problem is the heterogeneity of ICOs, although there were first approaches of standardization [16]. Additionally, recent research indicates that the ICO success heavily depends on its design parameters [9, 12, 17, 18]. Therefore, in-depth analysis of ICO design variations is necessary to better understand the phenomenon and react appropriately from an economic, societal, or regulatory side.

In particular, Information Systems (IS) research and specifically sociotechnical research needs to address this information technology driven phenomenon and provide a systematic understanding, as there is a need to investigate implications of the technology [19]. Classifying the extremely heterogeneous ICOs into predominant, lucid archetypes, analyzing them, and thereby getting a systematic understanding of the emergent phenomenon contributes to the current body of knowledge. Further, it allows to establish a common understanding of ICO designs, related consequences, and the application for investors and ventures. Yet, scientific research in the young research domain of ICOs is still scarce [5, 12, 18]. Boreiko and Sahdev [9] provide an overview of the evolution of ICOs. Chanson, Risius and Wortmann [5] compare ICOs to traditional crowdfunding mechanisms and Fridgen, Regner, Schweizer and Urbach [7] propose a taxonomy to classify ICO characteristics. They furthermore suggest four possible ICO archetypes as a basis for future research. However, in the rapidly evolving ICO landscape, enhancing the taxonomy [7] by adding additional cases might reveal necessary amendments to the taxonomy and further archetypes that occurred after November 2017. Although these research projects represent first important steps into the emerging domain, to date, a comprehensive and in-depth analysis of ICO archetypes is missing. Therefore, the goal of our research project is to bridge the existing gap by empirically investigating and analyzing ICO archetypes and evaluating ICOs in a structured manner. Therefore, we define the following research question: *Which quantitatively derived and qualitatively interpreted ICO design archetypes do exist, and which design parameters do differentiate them?*

To answer this question, we conduct a cluster analysis upon the refined ICO taxonomy of Fridgen, Regner, Schweizer and Urbach [7], to initiate the next step in ICO research. Compared to this existing study, we find more reliable results by increasing the clustering performance. We conduct a two-stage clustering approach, which yields in more accurate results, as the final clusters do not depend on a random selection of initial cluster centroids. [20-22]. By doing so, we aim to make a twofold contribution: First, we propose empirically based archetypes obtained from a sound clustering methodology providing a comprehensive understanding of the ICO phenomenon and of related implications for individuals as well as economic or regulatory organizations. Second, we aim to allow practitioners to conclude on concrete design suggestions for potential future ICOs with regard to the consequences arising from specific design decisions.

2 Research Method

In this section, we give an overview on our overall research approach, and resume with a detailed introduction of our cluster analysis. To identify meaningful archetypes of ICOs, we perform a cluster analysis, in line with IS literature and the exploratory research setting [23-25]. A cluster analysis is a statistical technique with the aim to group entities of similar kind into respective clusters. The variation within groups is minimized, whereas the variance between groups is maximized [20, 21]. In general, cluster analyses are applicable to describe generic archetypes of entities [21, 26]. In IS research, according to an analysis of 55 IS articles, researchers chose this method regularly to classify observations of specific objects of interest [27].

The cluster analysis follows three basic steps: First, we select the clustering variables. In chapter 3, after giving a general overview on blockchain, ICO, and design archetypes, we therefore review existing research on ICO classification, including the ICO design taxonomy of Fridgen, Regner, Schweizer and Urbach [7]. Second, we determine the appropriate cluster algorithm. Finally, we apply statistical methods to confirm the reliability as well as the validity of the results. We report the application of the hereinafter described research method in chapter 4. The qualitative interpretation of the archetypes and the conclusion follow in the remaining two chapters, 5 and 6.

Variables: The selection of clustering variables represents a fundamental step in cluster analysis because it highly affects the outcome [28]. Following a deductive approach [29], the chosen variables need to be closely linked to extant theory [22]. For this purpose, choosing a taxonomy's dimensions is a commonly applied approach [23]. Therefore, we use the 23 dimensions of the taxonomy from Fridgen, Regner, Schweizer and Urbach [7] as distinctive variables. Some researchers propose to perform a factor analysis as a pre-process and use the resulting factor scores for the clustering [28, 30]. However, literature does not recommend this approach if the data is not suitable for factor analysis due to dropping factors may then result in suboptimal clusters [22]. Furthermore, using factors hampers the interpretability of cluster outcomes [31, 32].

A*lgorithm:* After the selection of the cluster variables, we select an appropriate clustering algorithm. The application of hierarchical or non-hierarchical algorithms is well-recognized. However, both algorithms have various limitations when applied in an isolated way [22]. Hierarchical methods (e.g., Ward's algorithm) are highly sensitive to outliers [21, 33]. Non-hierarchical procedures require pre-specifying a number of clusters, which is difficult in an exploratory study field [27]. Therefore, instead of choosing one method, researchers developed two-stage clustering to improve the clustering performance and to receive more accurate results - combining the advantages of both methods [20, 22, 28, 34]. As this represents the expert consensus among IS researchers [27], we adopt this two-stage clustering process.

Validation: As a basis for valid clusters, Hair, Black, Babin and Anderson [21] suggest finding significant differences between the selected variables for the developed clusters. Thus, we use cross tabulation analysis to identify which variables significantly contribute to the differentiation of ICO archetypes [26]. Subsequently, we conduct post-hoc tests to compare single clusters.

Data sample: To provide a comprehensive perspective on ICOs in this paper, we collect a data sample consisting of 84 ICOs with each 23 categorical data points along the taxonomy's dimensions. For this purpose, and due to the lack of an exhaustive ICO database, we create an ICO longlist through the lists published by token information providers that are perceived as most reliable in the blockchain community, such as ICObench [35], Coindesk [13], and SmithAndCrown [36]. Our sample includes ICOs from different industries and from all over the world in the period spanning from January 2013 to July 2018. As ICOs are rarely restricted to national borders and even intermix existing industries, it is very difficult to quote reliable information on the geographical origin and industry assignment.

3 Foundations

3.1 Blockchain, Initial Coin Offerings, and Design Archetypes

Blockchain is a decentral data structure that allows to store transactions immutably, chronologically, and transparently in distributed networks. Recently, a blockchain use case called ICO has become a popular alternative financing method for organizations [6, 7, 9, 11, 37]. This phenomenon emerges due to the rise of the second generation of blockchain and the establishment of smart contracts. Smart contracts are referred to as computer programs that allow to implement business logic tamper-proof in blockchains [38]. This enables the development and execution of programs that invoke secure transactions between two or more parties with no need of knowing and trusting each other [3, 33]. As smart contracts are also able to control digital assets, they enable the issuance and distribution of digital tokens that reside on top of blockchains [39]. This mechanism to create and transfer tokens is the fundamental part of any ICO. The funds raised during an ICO typically finance blockchain-related projects [40]. In this way, an ICO represents an alternative to crowdfunding in venture financing [8]. A substantial difference to crowdfunding, however, is the tradability of tokens on secondary markets. Tokens do not necessarily entail ownership of a firm but can fulfil various functions [9]. For instance, they might act as a digital share in a project or grant access to a blockchain enabled platform [41].

Since the surge of the ICO phenomenon in 2017, there has been increasing academic attention spent to analyzing various aspects of ICOs. Empirically, Adhami, Giudici and Martinazzi [17] analyze the success determinants of ICOs, gathering financial data and looking at theoretically obtained input variables. Amsden and Schweizer [12] as well as Fisch [18] propose a different definition of success including also the token's listing status. Based on these studies, Boreiko and Sahdev [9] propose a further definition of success, distinguishing between top ICOs, average ICOs, and failed ICOs. Li and Mann [11] focus on how an ICO increases social welfare and discuss governance mechanisms of an ICO thereby proposing guidance to regulators.

Furthermore, first research steps to explore the underlying classification of ICOs have been made by Fridgen, Regner, Schweizer and Urbach [7]. They applied a structured in-depth analysis of ICOs to develop a taxonomy incorporating 23 dimensions. In their research outlook, they are already able to identify four basic ICO archetypes. However, as their study is limited to k-means clustering, their results strongly depend on the selection of initial cluster centroids [20-22]. Additionally, as the cluster analysis deals with the special case of categorical data, a more powerful distance measure should replace Euclidean distances [42]. Finally, as their focus is the development of a taxonomy, their cluster analysis remains a descriptive only first step towards the differentiation of ICO archetypes. Since recent research indicates that ICO design parameters significantly influence ICO success [9, 12, 17, 18], it is of vital importance to understand and analyze predominant ICO archetypes. Soh and Markus [43] and Trice and Beyer [44] argue that the empirical identification and evaluation of archetypes is a suitable method to create understanding about multifold and complex new phenomena. Existing research indicates that a selected meta-characteristic can classify and evaluate ICOs, however, these classifications in current research draw on a rather conceptual basis [7]. Thus, building upon empirically validated design parameters and applying an in-depth two-stage cluster analysis, our study represents an important step towards a better understanding of ICOs.

3.2 ICO Classification

Since blockchain is a dynamic and a very young research area, design parameters of ICOs are continuously evolving. Therefore, we undertake a critical reflection on the taxonomy of Fridgen, Regner, Schweizer and Urbach [7]. We find its meta-characteristic *Design parameters and characteristics of ICOs* applicable for our study as it comprehensively covers both, the purpose of the taxonomy as well as the purpose of the archetypes we aim to investigate. However, it generally is a valid limitation of taxonomies that additional cases can possibly not be classified within the existing dimensions. This is why Nickerson, Varshney and Muntermann [45] require a useful taxonomy to be extendible when new types of objects appear. The restriction that taxonomies are collectively exhaustive implies that a taxonomy is not final but needs to be extended incrementally by including additional dimensions and characteristics in the course of time. Thus, we follow the advice of Nickerson, Varshney and Muntermann [45], and revise the taxonomy of Fridgen, Regner, Schweizer and Urbach [7]. From the conceptual perspective, we deduct the dimensions on the basis of related literature and of semi-structured interviews with ICO practitioners. From the empirical perspective, we iteratively examine our data sample of 84 ICO cases and classify them into the taxonomy. We find that some new characteristics and dimensions appeared in the ICO environment, and thereby add or adapt dimensions and characteristics when necessary to cover all ICO cases. As a result, we suggest enriching the model by adding further characteristics to two existing dimensions and by re-defining the characteristics of three dimensions. Table 1 shows the final overarching dimensions forming our theoretical framework.

Table 1. Taxonomy of ICO design parameters based on Fridgen, Regner, Schweizer and Urbach [7]

<table>
<tr><th>Dimension</th><th colspan="12">Characteristics</th></tr>
<tr><td>Token implementation level</td><td colspan="4">on-chain</td><td colspan="4">native</td><td colspan="4">sidechain</td></tr>
<tr><td>Token purpose/type*</td><td colspan="2">usage</td><td colspan="2">work</td><td colspan="2">funding</td><td colspan="2">staking</td><td colspan="2">equity security</td><td colspan="2">non-equity security</td></tr>
<tr><td>Token supply growth</td><td colspan="4">fixed</td><td colspan="4">adaptive inflation</td><td colspan="4">fixed inflation</td></tr>
<tr><td>Token supply cap</td><td colspan="6">capped</td><td colspan="6">uncapped</td></tr>
<tr><td>Token burning</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Token distribution deferral</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Token holder voting rights</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Issuing legal structure</td><td colspan="6">foundation</td><td colspan="6">limited</td></tr>
<tr><td>Team company token share</td><td colspan="4">minority</td><td colspan="4">majority</td><td colspan="4">half</td></tr>
<tr><td>Team lockup period*</td><td colspan="4">no</td><td colspan="4">single period</td><td colspan="4">multiple periods</td></tr>
<tr><td>Pre-sale before ICO*</td><td colspan="3">no</td><td colspan="3">private</td><td colspan="3">public</td><td colspan="3">both</td></tr>
<tr><td>Pre-sale discount</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Planned occurrence</td><td colspan="4">multiple rounds</td><td colspan="4">single round</td><td colspan="4">unspecified</td></tr>
<tr><td>Registration needed</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Eligibility restrictions</td><td colspan="3">none</td><td colspan="3">geographic</td><td colspan="3">accreditation</td><td colspan="3">multiple</td></tr>
<tr><td>Purchase amount limit</td><td colspan="3">none</td><td colspan="3">minimum</td><td colspan="3">maximum</td><td colspan="3">both</td></tr>
<tr><td>Auction mechanism*</td><td colspan="6">yes</td><td colspan="6">no</td></tr>
<tr><td>Sales price</td><td colspan="6">fixed</td><td colspan="6">floating</td></tr>
<tr><td>Price fixing currency</td><td colspan="6">crypto</td><td colspan="6">fiat</td></tr>
<tr><td>Funding currency</td><td colspan="4">crypto</td><td colspan="4">both</td><td colspan="4">none</td></tr>
<tr><td>Funding cap*</td><td colspan="3">none</td><td colspan="3">hard cap</td><td colspan="3">soft cap</td><td colspan="3">multiple</td></tr>
<tr><td>Time horizon</td><td colspan="4">block time</td><td colspan="4">fixed date</td><td colspan="4">open end</td></tr>
<tr><td>Time-based discount</td><td colspan="4">none</td><td colspan="4">single rate</td><td colspan="4">multiple rates</td></tr>
<tr><td colspan="13">*Extensions & changes to the taxonomy of Fridgen, Regner, Schweizer and Urbach</td></tr>
</table>

Token purpose: To the current four token purposes, we add the two novel types *Equity security token* and *Non-equity security token*. Applying SEC regulation, a token represents a security if they meet all elements of the Howey test [46]. These include that the token embodies (i) an investment of money, (ii) in a common enterprise, (iii) with an expectation of profits [47]. An equity security token bears a dividend to the token holder, see for example the TAAS token. A non-equity security token behaves like a security but represents a loan for a specified time, and the founders are able to buy back the token, see for example the ZRCoin [48].

Team lockup period: Following IPO and Venture Capital literature, we apply the more common term *lockup* instead of vesting for the dimension, referring to the time window during which owners are not allowed to redeem their tokens [49].

Pre-sale before ICO: We add the characteristic *multiple* to the dimension, since the analysis of ICO cases reveals that some ICOs follow both, a private and public pre-sale.

Auction mechanism: Empirically, we observe the Dutch auction mechanism as the only implemented one so far, however different manifestations are possible. We therefore change this dimension's characteristics into *no* and *yes* thereby subsuming all kinds of theoretical auction mechanisms.

Funding cap: The analysis of ICO cases reveals that a specified soft cap does not necessarily trigger a remaining time limit of the ICO. It generally represents a minimum funding goal the team aims to raise in order to create a minimum viable product. Sometimes, analogous to the all-or-nothing mechanism in crowdfunding [50], if the ICO fails to reach the soft cap, the issuer returns all funds.

4 Cluster Analysis and Identification of ICO Archetypes

In this chapter, we apply the aforementioned research method and provide the quantitative results of the cluster analysis. We collect a data sample consisting of 84 ICOs with each 23 categorical data points along the taxonomy's dimensions. Since we select the taxonomy's dimensions as cluster variables, we need to avoid overweighting underlying constructs. This is an issue if the clustering variables are significantly correlated [22]. Therefore, we conduct a multiple correspondence analysis (MCA), which is as an extension of a principal component analysis for categorical data [51]. We obtain low eigenvalues of the resulting factors. This indicates that we should keep the initial 23 dimensions as clustering variables.

According to the chosen two-stage clustering process, the clustering algorithm starts with the hierarchical analysis. We apply Ward's method, which is the most commonly applied algorithm among the hierarchical methods [27] due to the production of reliable cluster results [23, 24, 52]. For the distance measure between categorical data points, literature recommends using the Jaccard, the Simple Matching, and the Dice coefficient [31, 53]. We test different measures and find that all produce highly similar results [23]. We then inspect the dendrogram and the scree-plot to determine the appropriate number of clusters [20]. This step reveals that five clusters represent the optimal number of clusters as any additional cluster would not significantly lower the total within cluster sum of squares. Additionally, we compute the average silhouette width and the gap statistic [54]. They both confirm the five-cluster solution. Next, we conduct non-hierarchical clustering with the results from the Ward's method as input to pre-specify the number of clusters. Among IS studies, researchers widely use the k-means approach with Euclidean distance measure [27]. However, research indicates that k-means is not the optimal approach to process categorical data since Euclidean distances are not meaningful on a discrete sample space [55]. Huang [56] therefore proposes a non-hierarchical clustering algorithm called k-modes, using a simple dissimilarity measure and substituting the means of the clusters with modes [33, 55]. The application of the k-modes algorithm to the dataset results in our five final clusters. Subsequently, we apply Pearson's χ^2 and Cramer's V, measures for the strength of a relationship, to analyze global differences across all clusters in the categorical data points [23, 24]. To compare single cluster differences, we use Pearson's χ^2 with correction for alpha inflation (Bonferonni style).

Table 2 provides an overview of the cluster analysis results. The results indicate the significant contribution of the taxonomy's ICO design dimensions chosen as cluster variables to the differentiation of ICO archetypes. The Chi-square reports significant values for most cluster variables, and the Cramer's V reports medium to strong association. The exceptions reflect some sales terms variables, i.e. the funding currency and the fixing of the price, closely related to the auction mechanism, as well as two time-related sales terms. Analysis reveals that the information gained from these variables is low, and there is low variation among clusters. We also conduct the clustering without these variables and received nearly identical results. Thus, in order to not lose information, we keep the variables in the taxonomy [45], as we perceive them as important dimensions in the overall characterization of ICOs.

Table 2. Results of cluster analysis

Dimension	Cluster					Significance tests		
	1 n=29	**2 n=20**	**3 n=19**	**4 n=9**	**5 n=7**	**χ^{2} [a]**	**V [b]**	**Pairwise post-hoc tests [c]**
Implementation level	onchain 93%	onchain 80%	onchain 84%	onchain 100%	native 86%	34.42 ***	0.453 **	1-5***;2-5***;3-5***
Purpose/ Type	usage 59%	usage 80%	usage 42%	usage 78%	staking 71%	27.73 *	0.287 *	
Supply growth	fixed 90%	fixed 80%	fixed 84%	fixed 89%	fix infl. 71%	47.87 ***	0.534 ***	1-5***;2-5***;3-5***;4-5**
Token supply cap	capped 97%	capped 90%	capped 89%	capped 100%	uncap. 100%	46.8 ***	0.746 ***	1-5***;2-5***;3-5***;4-5**
Token burning	no 72%	no 90%	no 58%	yes 89%	no 100%	22.64 ***	0.519 ***	1-4**;2-4**;4-5**
Distrib. deferral	yes 66%	no 70%	no 63%	yes 56%	yes 86%	10.95 *	0.361 **	
Holder voting rights	no 90%	no 90%	yes 63%	no 89%	yes 71%	27.53 ***	0.572 ***	1-3**;1-5**;2-3**2-5**
Issuing structure	limited 90%	limited 75%	limited 100%	found. 67%	found. 57%	23.69 ***	0.531 ***	1-4**;3-4**;3-5**
Team token share	minor. 97%	minor. 75%	minor. 100%	minor. 89%	minor. 100%	16.16 *	0.310 **	
Team lockup	single 59%	no 60%	multi. 47%	multi. 78%	no 57%	29.71 ***	0.421 **	1-2**;1-3*;1-4**;2-4**
Pre-sale before ICO	private 69%	no 70%	no 53%	public 56%	no 71%	44.01 ***	0.418 **	1-3***;1-2***;2-4**
Pre-sale discount	yes 100%	no 75%	no 79%	yes 78%	no 71%	43.86 ***	0.723 ***	1-2***;1-3***;1-5***;2-4*;3-4*

Registration needed	yes 93%	no 85%	yes 84%	yes 89%	no 86%	45.58 ***	0.737 ***	1-2***;1-5***;2-3***;2-4**;3-5**;4-5**
Eligibility restrictions	geogr. 55%	none 100%	geogr. 68%	none 56%	none 86%	51.34 ***	0.451 **	1-2***;1-4*;1-5*;2-3***;3-4**;3-5**
Planned occurrence	single 97%	single 50%	single 84%	single 89%	single 57%	20.07 *	0.346 **	1-2**
Purchase limit	none 72%	none 80%	none 79%	min. 44%	none 86%	23.91 *	0.308 **	1-4**;3-4**;2-4**
Sales price	fixed 86%	fixed 75%	fixed 89%	fixed 89%	fixed 57%	4.98	0.243 *	
Price fixing currency	crypto 55%	crypto 70%	fiat 68%	crypto 78%	crypto 100%	13.3 **	0.398 **	
Funding currency	crypto 83%	crypto 95%	crypto 63%	crypto 67%	crypto 100%	9.36	0.334 **	
Funding cap	hard 66%	hard 45%	multi. 74%	multi. 67%	none 71%	41.56 ***	0.406 **	1-3**;1-5**;4-5**
Time horizon	fixed 90%	fixed 70%	fixed 95%	fixed 89%	fixed 71%	8.59	0.226 *	
Auction mechanism	no 97%	no 90%	no 100%	no 100%	no 71%	9.08	0.329 **	
Time based discount	multiple 52%	multiple 55%	no 58%	no 56%	multiple 43%	14.09	0.290 *	

* p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001
a Percentages in one cluster which show a given characteristic
b Threshold ***V >=0.5; **V>=0.3; *V>=0.2
c Post hoc significances between single clusters are tested using Pearson′s χ2

5 Analysis and Implications of ICO Archetypes

The cluster analysis grouped five distinct entities of similar kind with regard to the respective ICO's design characteristics, minimizing the variance within the groups. Due to the initial hierarchical clustering approach, which does not require to pre-specify a number of clusters, our analysis yields in a different number of clusters compared to the archetypes from Fridgen, Regner, Schweizer and Urbach [7]. Additionally, since our dataset also includes novel forms of ICOs, our five clusters differentiate more particularly with regard to the token terms including the token purpose. Each of these clusters thereby form a unique archetype which we investigate in the following.

Archetype 1: The average ICO. This ICO archetype represents the largest cluster. We perceive its characteristics as the most typical ones since it resembles the patterns of a traditional crowdfunding campaign. Based on top of an existing blockchain, the issuer raises a capped amount of funding. Capping the amount possibly avoids being perceived greedy and may mitigate the risk of regulatory attention [57]. A private pre-sale allows the issuer to raise money prior to the regular sale. The team can then focus on developing the product early, whereas the early investors benefit from a discount. This archetype implements a usage token providing access to a service or platform and does not transfer voting rights or company shares to the token holders. It therefore tends to target investors who are interested in the actual use case, i.e. the access to a service or platform provided, rather than e.g. investment returns.

Archetype 2: The liberal ICO. This archetype shows comparably less governance from issuers with regard to sales terms and issuer terms. It tries to maximize the target group of prospective buyers, since it does not require prior registration. Furthermore, it does not impose geographic restrictions nor restrict the access to accredited investors. Additionally, this archetype does not offer any pre-sale and there is no purchase amount limit. This indicates that the tokens are sold on a first-come, first-served basis without favoring wealthy or institutional investors. Thereby, we consider that this archetype corresponds the most to the truly global and inclusive blockchain idea [37]. This archetype partially includes those ICOs planning multiple funding rounds instead of a single round only. In venture capital, funding traditionally takes place in multiple rounds, one consequence is that the issuing team remains incentivized [58]. This is why blockchain experts also believe that an iterative funding approach could be the future of ICOs [59].

Archetype 3: The visionary ICO. In many of its design parameters, this ICO archetype offers several value propositions. The issuer grants voting rights to its investors which can thereby participate in the initiative's development. Additionally, the archetype sets lockup periods for the token share allocated to the issuer. These lockups prevent the team from selling their tokens directly after the closing of the ICO, which stabilizes the post-ICO token price [60]. Further, this archetype specifies both, a soft and a hard cap for the ICO. The announcement of a clear funding target range conveys the message that the issuer intends to raise an amount aligned with the expected costs of network development [57]. In many cases, the whitepaper specifies that all funds are returned to investors if the ICO fails to reach the soft cap [12]. This procedure reduces the investor's risks and indicates that the team links its funding tightly to the development costs. Thus, we conclude that this ICO archetype goes beyond being just a funding mechanism, but targets at investors that truly believe in the business model and its long-term success.

Archetype 4: The compliant ICO. The prevailing pattern in this archetype represents the regulatory orientation of the ICO design. By burning the unsold tokens post-ICO, the issuer keeps the percentages in token allocation between issuer and investors stable. Usually, the token burning benefits the token holders, since it decreases the total number of available tokens, and, consequently, may increase the value of each individual token [61]. However, a controlled appreciation of the token value may attract regulatory attention, since then the token could be considered as a security [62]. Another peculiarity of this archetype is the structuring of the issuing legal entity as foundation. Recently, there has been a trend in the ICO universe to divide the corporate structure into two separate entities, where a foundation runs the ICO, and another entity runs the business operations [10]. This enables the legal separation of the liabilities associated with the ICO. With regard to the sales terms, the issuer has more information and control over the investors as they need to register before they can purchase tokens. Additionally, pre-defined purchase limits restrict the token sale. Limiting the maximum purchase amount can enhance a wider distribution of the tokens, thereby preventing a token concentration of a single investor. A concentrated token share distribution could raise regulatory issues regarding secondary market trading. Thereby, we conclude that the design of this ICO archetype, more than others, takes into account the current regulatory uncertainty and seeks to comply with potential upcoming ICO regulation.

Archetype 5: The native ICO. Differences regarding the technical token terms predominantly characterize this archetype. In particular, the token implementation level represents a striking characteristic of this archetype. Whereas many tokens use the ERC20 token standard from the Ethereum blockchain, this archetype, however, distributes tokens that are native to their own blockchain. These tokens are often referred to as protocol tokens. They may be used as simple currency or might have other use cases, such as a stake to participate in a network. Often, the developers aim to create novel use cases based on these tokens. These innovative features appear to aim at overcoming challenges of existing blockchain solutions such as scalability [39]. Another unique characteristic in this cluster is the uncapped supply of tokens, so all investors are able to buy as many tokens as they desire. We conclude that this archetype comes with interesting specificities especially for blockchain enthusiasts.

Summarizing, we learn that the five archetypes differ from each other with regard to value propositions, target groups, and existing challenges. From an ICO issuer perspective, a key task constitutes the definition of a clear value proposition. This ultimately translates into the respective target group of investors. For example, designing an ICO similar to archetype 3, the visionary ICO, might also attract investors interested and engaged in the further development of the network. Many ICOs incorporate a liberal design, i.e. archetype 2, corresponding to the fundamental idea of the blockchain technology. Implementing a liberal ICO design, however, one might end up having investors exploiting the non-existing restrictions (e.g. money laundering purposes). From an investor perspective, it is of vital importance to know what the objectives of the ICO issuer are to better understand the token prize development. Being interested in the long-term vision, it might make sense to look out for ICOs with designs similar to the archetypes 1 and 3, the average and the visionary ICO. If an investor primarily seeks a promising financial return, investing into compliant ICOs as archetype 4 might be the right way. In that case, the ICO might attract higher regulatory attention due to the token burning that leads to a potential reduction of the investor's risk. Taking the amount of cases within each cluster into account, we observe that most ICOs currently do not consider regulatory issues. This may be due to the novelty of the phenomenon. Partly, the global nature of blockchain applications may make it difficult to consider the regulatory variety across countries. ICO issuers might therefore decide to ignore any regulatory aspects so far. This picture, however, is likely to change since the ICO phenomenon is attracting more attention recently, especially from the regulatory side.

6 Conclusion and Outlook

ICO as a novel funding mechanism represents a promising example of a blockchain use case that recently draws attention in both, research and practice. Although first research projects analyzed specific aspects of the emerging phenomenon, we poorly understand the implications of ICOs yet. Thus, in this research paper, we bridged this research gap and investigated ICOs with regard to their design parameters and focused on the identification as well as qualitative analysis of predominant archetypes. To do so, we first enriched the established taxonomy for ICO design [7] to account for recent developments in the fast evolving blockchain domain. Second, we used the taxonomy's 23 dimensions as clustering variables and conducted a cluster analysis on 84 ICO cases. As a result, we proposed five ICO archetypes which illustrate different combinations and dominant aspects within the ICO design parameters. Further, we examined these clusters and presented a qualitative interpretation for each archetype.

Before emphasizing our contributions to both research and practice, we acknowledge some limitations as well as highlight promising starting points for future research. First, we limited our sampling procedure to ICOs with exhaustive data available to allow for comprehensive structuring according to the taxonomy's dimensions. Second, we used a convenient data sample, which represents a representative share of the total ICO market only. Third, this paper only addresses ICO design parameters, rather than other ICO aspects, which have been examined in traditional crowdfunding literature, such as the business model and industry or the quality of marketing. These aspects should be subject to further research that might help to better understand the ICO phenomenon. Fourth, the ICO market is highly dynamic and most ICO issuers are startups. Thereby, token sale models are constantly evolving, leading to dynamic emergences of novel ICO design patterns. However, we strongly believe that our empirically obtained archetypes comprehensively describe the current ICO market. Finally, it also remains for future research to investigate how the fast developments of blockchain technology influences the future of ICOs.

The theoretical contributions of our research address the research gap in three ways: First, we provide a systematic and comprehensive overview on predominant ICO designs. We suggest five ICO archetypes with different value propositions, target groups, and challenges. The better fitting clustering method and the qualitative discussion and interpretation of the archetypes allow to abstract from single peculiarities of specific ICOs and enable thereby generalizable propositions. We therefore systemize the findings generated by Fridgen, Regner, Schweizer and Urbach [7]. Second, the archetypes extend existing ICO classifications by various aspects and allow for generalizable findings, instead of taking into account single characteristics. Third, we lay the foundation for further research in the area of ICOs. Since the archetypes are theoretically grounded on an existing taxonomy and empirically verified, they provide a more systematic and in-depth perspective on the phenomenon. This will help synthesize research on ICOs and identity promising research avenues.

Besides our theoretical contributions, our research provides practitioners with various backgrounds and perspectives on the ICO phenomenon. First, the classification into predominant archetypes may provide structured guidance for ventures that plan to conduct an ICO. Second, from an investor point of view, the archetypes can lead to more informed and grounded investment decisions. Third, for traditional financial intermediaries, including early stage venture capitalists or crowdfunding platforms, the enriched taxonomy and archetypes may help to characterize potential competitors. Fourth, our approach to structure the heterogeneous ICO market through design archetypes allows to reduce complexity, which may help regulators to perform regulatory tasks more effectively. This ultimately reduces the uncertainty in the market for all participants.

References

1. Bharadwaj, A., El Sawy, O., Pavlou, P., Venkatraman, N.: Digital business strategy: toward a next generation of insights. MIS Quarterly 37, 471-482 (2013)
2. Gimpel, H., Hosseini, S., Huber, R., Probst, L., Röglinger, M., Faisst, U.: Structuring Digital Transformation: A Framework of Action Fields and its Application at ZEISS. Journal of Information Technology Theory and Application 19, 31-54 (2018)
3. Glaser, F.: Pervasive Decentralisation of Digital Infrastructures: A Framework for Blockchain enabled System and Use Case Analysis. In: Proceedings of the 50th Hawaii International Conference on System Sciences. Waikoloa Village, Hawaii (2017)
4. Catalini, C., Gans, J.S.: Initial Coin Offerings and the Value of Crypto Tokens (2018)
5. Chanson, M., Risius, M., Wortmann, F.: Initial Coin Offerings (ICOs): An Introduction to the Novel Funding Mechanism Based on Blockchain Technology. In: Proceedings of the 24th Americas Conference on Information Systems. New Orleans (2018)
6. Schweizer, A., Schlatt, V., Urbach, N., Fridgen, G.: Unchaining Social Businesses: Blockchain as the Basic Technology of a Crowdlending Platform. In: Proceedings of the 38th International Conference on Information Systems. Seoul (2017)
7. Fridgen, G., Regner, F., Schweizer, A., Urbach, N.: Don't Slip on the Initial Coin Offering (ICO) - A Taxonomy for a Blockchain-enabled Form of Crowdfunding. In: Proceedings of the 26th European Conference on Information Systems. Portsmouth (2018)
8. Ravikant, Naval, https://genius.com/Naval-ravikant-the-bitcoin-model-for-crowdfunding-annotated, 11.07.2018
9. Boreiko, D., Sahdev, N.K.: To ICO or not to ICO – Empirical Analysis of Initial Coin Offerings and Token Sales (2018)
10. Zetzsche, D.A., Buckley, R.P., Arner, D.W., Föhr, L.: The ICO Gold Rush: It's a Scam, It's a Bubble, It's a Super Challenge for Regulators (2018)
11. Li, J., Mann, W.: Initial Coin Offering and Platform Building. In: 2018 Western Finance Association Meeting / Conference. Coronado (2018)
12. Amsden, R., Schweizer, D.: Are Blockchain Crowdsales the New 'Gold Rush'? Success Determinants of Initial Coin Offerings (2018)
13. Coindesk, https://www.coindesk.com/ico-tracker/, 01.08.2018
14. Vigna, P., Shifflett, S., Ostroff, C.: What Crypto Downturn? ICO Fundraising Surges in 2018. Wall Street Journal (2018)
15. Clayton, J.: Statement on Cryptocurrencies and Initial Coin Offerings. U.S. SEC (2017)
16. Estonian Financial Supervision Authority, https://www.fi.ee/index.php?id=21662, 14.08.2018
17. Adhami, S., Giudici, G., Martinazzi, S.: Why do businesses go crypto? An empirical analysis of initial coin offerings. Journal of Economics and Business (Forthcoming) (2018)

18. Fisch, C.: Initial Coin Offerings (ICOs) to Finance New Ventures: An Exploratory Study (2018)
19. Beck, R., Avital, M., Rossi, M., Thatcher, J.B.: Blockchain Technology in Business and Information Systems Research. Business & Information Systems Engineering 59, 381-384 (2017)
20. Aldenderfer, M.S., Blashfield, R.K.: Cluster analysis. Sage Publication, Beverly Hills (1984)
21. Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E.: Multivariate Data Analysis. Pearson Education, Upper Saddle River (NJ) (2013)
22. Ketchen, D.J., Shook, C.L.: The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal 17, 441-458 (1996)
23. Haas, P., Blohm, I., Leimeister, J.M.: An Empirical Taxonomy of Crowdfunding Intermediaries. In: Proceedings of the 35th International Conference on Information Systems. Auckland (2014)
24. Malhotra, A., Gosain, S., El Sawy, O.A.: Absorptive Capacity Configurations in Supply Chains: Gearing for Partner-Enabled Market Knowledge Creation. MIS Quarterly 29, 145-187 (2005)
25. Püschel, L., Schlott, H., Röglinger, M.: What's in a Smart Thing? Development of a Multi-layer Taxonomy. In: 37th International Conference on Information Systems. Dublin (2016)
26. Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Cluster Analysis. Wiley, Chichester (2011)
27. Balijepally, V., Mangalaraj, G., Iyengar, K.: Are We Wielding this Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research. Journal of the Association for Information Systems 12, 375-413 (2011)
28. Punj, G., Stewart, D.W.: Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research 20, 134-148 (1983)
29. Ketchen, D.J., Thomas, J.B., Snow, C.C.: Organizational Configurations and Performance: A Comparison of Theoretical Approaches. Academy of Management Journal 36, 1278-1313 (1993)
30. Ding, C., He, X.: K-means clustering via principal component analysis. In: Proceedings of the 21st International Conference on Machine Learning, p. 29. ACM, Banff (2004)
31. Berkhin, P.: A Survey of Clustering Data Mining Techniques. In: Kogan, J., Nicholas, C., Teboulle, M. (eds.) Grouping Multidimensional Data, pp. 25-71. Springer, Berlin, Heidelberg (2006)
32. Carroll, J.D., Green, P.E., Schaffer, C.M.: Interpoint Distance Comparisons in Correspondence Analysis. Journal of Marketing Research 23, 271-280 (1986)
33. Xu, X., Weber, I., Staples, M., Zhu, L., Bosch, J., Bass, L., Pautasso, C., Rimba, P.: A Taxonomy of Blockchain-Based Systems for Architecture Design. In: 2017 IEEE International Conference on Software Architecture, pp. 243-252. IEEE, Gothenburg (2017)
34. Milligan, G.W., Sokol, L.M.: A Two-Stage Clustering Algorithm with Robust Recovery Characteristics. Educational and Psychological Measurement 40, 755-759 (1980)
35. ICObench, https://icobench.com/, 01.08.2018
36. Smith and Crown, https://www.smithandcrown.com/, 01.08.2018
37. Porru, S., Pinna, A., Marchesi, M., Tonelli, R.: Blockchain-Oriented Software Engineering: Challenges and New Directions. In: Proceedings of the 39th International Conference on Software Engineering Companion. IEEE Press, Buenos Aires (2017)
38. Szabo, N.: Smart Contracts: Formalizing and Securing Relationships on Public Networks. First Monday (1997)
39. Buterin, V.: A next-generation smart contract and decentralized application platform. White Paper (2014)

40. Chuen Kuo, D.L., Guo, L., Wang, Y.: Cryptocurrency: A New Investment Opportunity? The Journal of Alternative Investmens 20, 16-40 (2018)
41. Ehrsam, F.: Blockchain Tokens and the dawn of the Decentralized Business Model. The Coinbase Blog (2016)
42. Boriah, S., Chandola, V., Kumar, V.: Similarity Measures for Categorical Data: A Comparative Evaluation. In: Proceedings of the 2008 SIAM International Conference on Data Mining, pp. 243-254. Society for Industrial and Applied Mathematics, Atlanta (2008)
43. Soh, C., Markus, M.L.: Business-to-Business E-Marketplaces: A Strategic Archetypes Approach. In: Proceedings of the 23rd International Conference on Information Systems. Barcelona (2002)
44. Trice, H.M., Beyer, J.M.: Studying organizational cultures through rites and ceremonials. Academy of Management Review 9, 653-669 (1984)
45. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. European Journal of Information Systems 22, 336-359 (2013)
46. Bramanathan, R.: Introducing the Blockchain Token Securities Law Framework. The Coinbase Blog (2016)
47. Tsigulev, V.: Howey test: when a token can be treated as a security. Principles for blockchain tokens development and sale. Medium (2017)
48. Lielacher, Alex, https://cryptonews.com/exclusives/how-to-choose-the-right-security-tokens-to-invest-in-1871.htm, 10.08.2018
49. Smith, J.K., Smith, R.L., Bliss, R.: Entrepreneurial Finance. Stanford University Press (2011)
50. Cumming, D., Leboeuf, G., Schwienbacher, A.: Crowdfunding Models: Keep-It-All vs. All-Or-Nothing (2015)
51. Husson, F., Josse, J., Pages, J.: Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data? Applied Mathematics Department (2010)
52. Van de Vrande, V., De Jong, J.P., Vanhaverbeke, W., De Rochemont, M.: Open innovation in SMEs: Trends, motives and management challenges. Technovation 29, 423-437 (2009)
53. Finch, H.: Comparison of Distance Measures in Cluster Analysis with Dichotomous Data. Journal of Data Science 3, 85-100 (2005)
54. Tibshirani, R., Walther, G., Hastie, T.: Estimating the Number of Clusters in a Data Set Via the Gap Statistic. Journal of the Royal Statistical Society Series B (Statistical Methodology) 63, 411-423 (2001)
55. Chaturvedi, A., Green, P.E., Caroll, J.D.: K-modes clustering. Journal of Classification 18, 35-55 (2001)
56. Huang, Z.: Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery 2, 283-304 (1998)
57. Buterin, Vitalik, https://vitalik.ca/general/2017/06/09/sales.html, 07.07.2018
58. Davila, A., Foster, G., Gupta, M.: Venture capital financing and the growth of startup firms. Journal of Business Venturing 18, 689-708 (2003)
59. Petrovcic, D.: ICOs, don't bite off more than you can chew. hackernoon.com (2017)
60. Lee, S.: Three Reasons Why The ICO Market Is Cooling Down, But Crypto And Blockchain Are Here To Stay. Forbes (2017)
61. Ferrara, N.: 'Token Burning' And Other Crypto Jargon Simplified. Forbes (2017)
62. Ledbetter, A.D., Dykes, T.C.: SEC eats away at Munchee "utility tokens": guidance for ICOs. DLA Piper (2017)

Capturing Value from Data: Exploring Factors Influencing Revenue Model Design for Data-Driven Services

Tobias Enders[1], Ronny Schüritz[1], and Wiebke Frey[1]

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Karlsruhe, Germany
{tobias.enders,ronny.schueritz}@kit.edu
{wiebke.frey}@student.kit.edu

Abstract. In recent years organizations have started utilizing big data and advanced analytics not just to support decision-making to raise process efficiencies, but also to engage in new data-driven services. These data-driven services complement the current product and service portfolio and create additional value for customers. In order to capture the value created, organizations need to design sustainable revenue models consisting of a revenue (how) and pricing (how much) mechanism. In order to develop a deeper understanding of one part of the decision-making process on revenue models, we apply a qualitative study and analyze the results through the lens of rational choice theory. Based on the interviews, we derived four factors – service characteristics, provider interests, customer interests, and market factors - influencing the design. By this, we contribute to the general understanding of the design of revenue models and enable further investigation into this field of research.

Keywords: Big Data, Business Model, Revenue Model, Revenue Mechanism

1 Introduction

Every organization is concerned by the dawning era of big data analytics [1]. They are pushed to make use of data analytics to stay innovative and ahead of competitors [2]. Organizations mostly leverage data analytics to support their decision-making and improve process efficiencies, e.g. by continuously monitoring market trends to better allocate their sales force across different business units. However, in order to stay competitive in an ever servitizing economy, companies may use data analytics to create new products and services [3]. This can be achieved by either wrapping the product or service with data analytics [4] or by introducing entirely new data-driven business models [5]. Rolls-Royce, for example, has gone down the path of reinventing its service portfolio by enriching the use of a physical product – jet engines - with data analytics. By providing detailed product usage data (e.g. fuel consumption, temperatures, altitude), its customers can drive efficiencies in product usage [6].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Data-driven services are characterized by relying on data as their key resource and analytics to provide value to the customer. To take advantage of the value creation opportunities of these data-driven services, companies must design a suitable revenue model as part of their overarching business model [7–9] (cf. Figure 1). A process that is found to be a major challenge for organizations that approach datatization [10]. The question of how to design revenue models has already been challenging in the context of servitization and has not yet been answered conclusively [11]. Initial studies shed light on the availability and usage of revenue models for data-driven services in general [e.g. 11, 12]. Companies still need to decide which model is appropriate for their particular data-driven service. We want to investigate this decision-making process to approach this rather underappreciated field in research. In this explorative study we first investigate the decision-making process to build a better understanding of the factors influencing the choice of a revenue model. By that we aim to contribute to the general understanding of revenue models in the context of data-driven services. In particular, we contribute factors that represent an influence on the decision-making process for revenue mechanisms and we demonstrate that rational choice theory can be leveraged to explain our observations.

We initially focus on the analysis of data-driven services due to the current interest and demand in the market. In contrast to other services (e.g. air travel, maintenance services), data-driven services are characterized by being scalable through IT infrastructure and the fact that data, as a key resource, can be reused simultaneously at no additional cost [14]. Our research is guided by the overarching research question:

"What are the factors influencing the decision-making process for a revenue mechanism for data-driven services in a B2B environment?"

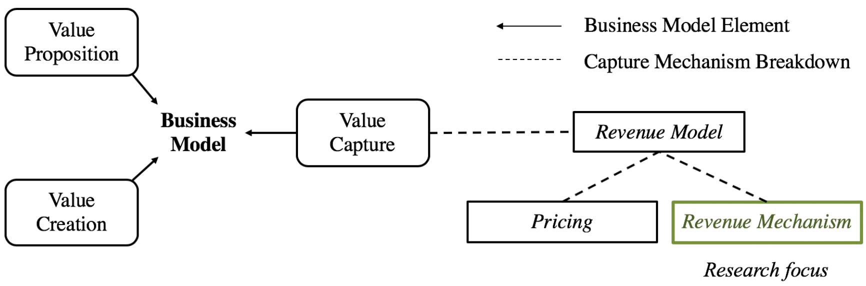


Figure 1. Research focus

We conducted interviews with 14 experts that were involved in the design of revenue models for data-driven services. By applying a qualitative content analysis and looking at the results through the lens of rational choice theory, we derived and analyzed four factors influencing the decision-making of the revenue mechanism for a data-driven service.

The remainder of the paper is structured as follows: section 2 introduces data-driven offerings and elaborates on the current state of research on revenue models. Section 3 illustrates how we collected data and the means of analyzing it. The results of our interview study are being presented in section 4. The paper closes with a brief summary, discussion, study limitations and managerial implications in section 5.

2 Related Work

In this section, we give a brief introduction to data-driven offerings and revenue models. Furthermore, we show that there is a need to develop an understanding of the factors that influence the choice of a revenue model.

2.1 Data-Driven Offerings

With the ever-increasing amount of data being generated, many organizations have realized that there is value hidden within this data. Often, data is generated as a byproduct of other activities e.g. through the use of sensors that connect the digital and physical world [15, 16]. An increasing amount of enterprises are making use of that data by developing new data-driven services or shifting their entire business model to be more data-driven. For instance, machine manufacturers start collecting real-time product usage data in order to move from reactive to predictive maintenance activities to reduce the costs of downtime [17].

Exploring the field of big data and its applications, e.g. for developing data-driven offerings, is not new. For instance, Manyika et al. [16] provide a broad overview on how the use of big data and services impact innovation, competition, and productivity. Chen et al. [15] explore ways in which insights derived from the use of big data can have an impact on different domains while highlighting the importance of the evolution of business intelligence and analytics over the past decades. Hartmann et al. [5] develop a taxonomy of data-driven business models and argue that the effective use of big data, e.g. by offering purely data-driven services, could lead to a competitive advantage. In their recent work, Wixom and Ross [4] propose three ways for enterprises to monetize their data: internal improvements, wrapping, and selling. They argue that besides deriving internal efficiencies, organizations can engage in becoming an information business by using data analytics to offer stand-alone services (selling) or to enhance the value of an existing core product or service through analytics (wrapping).

As organizations start developing new data-driven offerings, the question on how to capture the value of such offerings emerges and organizations are still struggling with answering it [10]. Revenue models describe the mechanism to capture this value and function as a critical component of the overall business model of the organization [18].

2.2 Revenue Models for Data-Driven Offerings

Literature on revenue models offers a multitude of perspectives on what they are: “revenue streams” [19] “pricing mechanisms” [20] and “payment model” [21] are some of the most common ones. While some of them have a customer view (e.g. payment model), others focus on the way the service provider captures an offering’s value (e.g.

revenue streams). Eventually, the meaning of a revenue model boils down to being a mechanism to capture the value of a good or service [8]. The definition of a sustainable revenue model is crucial for the viability of a business. Therefore, scholars agree that the revenue model is a critical component of the overall business model [7, 18, 22, 23].

The revenue model describes an overall architecture, which essentially consists of two key elements: revenue mechanism and price setting [8]. While price setting deals with how much to charge for a service or product, the revenue mechanism targets the question how to charge. The process of pricing a new product or service usually follows one of three approaches: cost-based, competition-based or value-based pricing. Since pricing is a complex field of research by itself, our study focuses on the decision-making process of revenue mechanisms. A study conducted by Schüritz et al. [12] provides an initial overview of the revenue mechanisms most commonly used by organizations offering data-driven services: in the availability-based model, also called subscription model, customers pay a service fee for a particular time period in which the service is being made available – independent of how much the offering is essentially used during that period [24, 25]. In contrast, a usage-based revenue mechanism only allows the provider to earn revenue when the service is actually used by the customer. This requires the definition of a unit of measure for the service to be charged [26]. Another model is so-called performance-based; in this case, the compensation of the provider is dependent on outcome generated for the customer [27, 28]. Also, multi-sided revenue mechanisms exist; they involve two or more interdependent customer groups. Google Adverts is a popular example of this model since it acts as a 'hub' that connects an advertiser and an end customer indirectly.

Data-driven services may also be used as a free add-on to an existing product or service. In such cases, the organization has chosen an indirect payoff model since the service is paid for by the revenue stream generated through the core offering itself.

Previous studies, such as the one conducted by Schüritz et al. [12], have shown that organizations have deployed various revenue mechanisms. However, little is known about the decision-making process that lets an organization select a revenue mechanism and reject others. A recent study conducted by Sprenger et al. [13] sheds a light on this topic by exploring the evolutionary changes of revenue mechanisms for digital offerings to support managerial decisions. The scholars suggest that the choice of the most suitable revenue mechanism depends on (1) the type of digital offering, (2) the stage of evolution and (3) six additional constrains. However, most research conducted on the topic of revenue models for digital businesses focuses on pricing decisions [29] or explores the trade-off between free vs. paid approaches for digital content [30].

While the work conducted by Schüritz et al. [12] and Sprenger et al. [13] sheds some light on revenue mechanisms for data-driven services in general, no conclusive work has been done to understand the decision-making process on revenue mechanisms.

Picking a revenue mechanism is a decision made on an organizational level. Nevertheless, it always comes down to an individual or group of individuals deciding. The process of determining which options are available, then choosing the most preferred one is widely discussed in rational choice theory. Individuals aim to maximize their utility (u), which can be understood as the organization's objective such as profit or revenue maximization. Since employees are incentivized to contribute to the firm's objective, e.g. through bonus payments, they adopt this view in the decision-making process. Choosing among alternatives A = {a1, a2,…,aj} is further guided by personal preferences and constrains, which may help us in understanding how decisions are made [31].

3 Research Design

To get a better understanding of the factors that influence the choice of a particular revenue mechanism for data-driven services, we follow a rigorous qualitative research process that is based on a set of interviews. Despite the fact that this study leverages procedures strongly associated with the grounded theory methodology (GTM) (e.g. open and axial coding), it omits elements such as memoing and theoretical sampling [32]. Therefore, we consider our approach for data collection and analysis a qualitative content analysis based on Krippendorff [33] and Bengtsson [34].

3.1 Data Sources

In order to understand how providers of data-driven services have designed their revenue model, we needed to collect data that gives us precise information on all the influencing factors, which were considered during the decision-making process.

In a first step, we collected data of publicly available cases from service providers that offer data-driven services in a B2B context (e.g. through websites, customer references, and news articles). Our working assumption is that the results of this study may differ when looking at B2B vs. B2B services. For this reason and limitations in resources available, we initially focus on one type only. The identified use cases served as a pre-study to inform next steps of our research. In particular, we were able to define more specific criteria for the sampling of interviewees and for the development of tailored questionnaires. Since the data collected through the case analysis did not provide enough detail to derive reasons why particular revenue models have been chosen, we decided to collect data by conducting a series of interviews with representatives of service providers that offer data-driven services and that were involved in the decision-making process.

Our prior collection of service providers of data-driven services served as sample frame for our interview study. All of these companies have already implemented data-driven services and have gone through the decision-making process of selecting a revenue mechanism. For the sampling of the interviewees, we followed a criterion-i purposeful sampling approach [35]. The criteria were defined that only representatives, who were directly involved in the decision-making process are subject of interest for interviews. We therefore focused on approaching the product managers and heads of service for the particular data-driven service via LinkedIn. The expected small number of respondents were either themselves available for an interview or forwarded our request to a colleague with better insight in the design of the revenue model. This approach has yielded a total of 14 interviews, which have been conducted in the time frame between August 2016 and June 2017. The interviewees hold various positions within their organizations such as product manager, head of service and head of innovation. The focused data-driven services are offered by global leaders in the fields of manufacturing, IT services, logistics and telecommunications. An overview of the interviewed organizations partners is given in Table 1.

Table 1. Overview of interviewed companies

Company	*Revenue (2017 in bn €)*	*Employees*	*Countries active*	*Number of interviews*
Manufacturing I	3,8	11.000	50	1
Manufacturing II	84,3	370.000	>150	1
Manufacturing III	0,8	3.500	40	1
Manufacturing IV	78,1	400.000	60	1
Manufacturing V	4,2	19.000	60	1
Manufacturing VI	3,6	13.500	70	1
Manufacturing VII	43,0	160.000	79	1
Manufacturing VIII	8,7	61.000	>100	1
Teleco	52,0	130.000	24	2
IT Services	1,0	7.000	7	2
Logistics[1]	164,3	290.000	>150	2

In order to create an open discussion situation in which the interviewee is willing to discuss the topic, a semi-structured interview approach is chosen. The questions asked during the interviews covered four themes: motivation to introduce a data-driven service, service type offered, revenue mechanism selection, and pricing strategy. While all of the sections contributed to the results of the study, the focus lay on the third one: revenue mechanism. Interviewees were asked why they selected a particular model and why others were rejected. The interviews took place in two phases: the first set of eight interviews were used to explore the inquiry and inform phase two, allowing for a set of more in-depth questions. In the subsequent set of interviews, the same questions as in phase one were asked and, additionally, some assumptions derived from the initial interviews were explored. Most of the interviews have been conducted over the phone with few exceptions where in-person meetings were made possible. Interviews have been transcribed except for one case in which the interviewee preferred the researcher to take notes.

3.2 Data Analysis

The 14 interviews are analyzed using qualitative content analysis. In this particular case, two coding cycles are conducted: to account for the explorative nature of topic, we pick an open coding approach for the first iteration, followed by axial coding [36]. The software MAXQDA is used to support this work.

[1] Figures refer to the parent company of the organization interviewed

In the first iteration, we start with no pre-defined list of codes. The transcribed interviews are labelled according to our research objective, identifying all factors that are influencing the choice of the revenue mechanism. Initially, 32 different codes are identified by the researchers. In a subsequent step, codes are grouped into categories and subcategories by two researchers in a workshop. The outcome of the coding of the first eight interviews serves as an input for the next interview set.

For the second iteration, we choose an axial coding approach. This step ensures that the codes identified in the first iteration are reassembled and that categories and sub-categories relate to each other accordingly [37]. Glaser [38] stresses the importance of this step since it requires sharpening the code for achieving its best fit. The final number of interviews we conducted was driven by our sampling criteria and by the availability of the interviewees during that time. Furthermore, we did see a saturation in the data in the final interviews.

Finally, to verify the results, we ask two independent researchers to code all interviews again based on the coding structure derived from the second coding cycle. To ensure objectivity and validity of the results, we calculate the intercoder reliability as an indicator of measurement consistency. An 85% mapping of the coded segments ensures a high confidence level in the results. Discrepancies between the researchers are discussed until an agreement is reached.

4 Influencing Factors

Designing a revenue model is a critical part of releasing new offerings to market as it describes the process of capturing the value of the offering. In some cases, the revenue model may even decide if the offering becomes a success [7]. Therefore, the decision for a particular revenue model is of high importance. Based on our interviews, we understand that there are two key decisions the provider has to make: how (revenue mechanism) and how much (price) to charge the customer. Unlike for the design of a product or service, we could not see formal processes or methods in the organization to decide on a revenue model, but we identified a series of factors that influence the decision-making process. When looking at an organization as a decision-making unit, the factors can be regarded through the lens of the rational choice theory (RCT) as preferences and constrains that inform the decision-making process. RCT postulates that the agent (an individual or organization) aims to maximizing the outcome of a rational decision while having the choice among multiple alternatives. The maximization of the net benefit, or utility, is driven by the benefit and costs as well as the level of risk that arises [31]. While the service provider eventually is the one to make the decision on the shaping of the revenue mechanism, decision-making on the customer side (i.e. purchase vs. non-purchase of a service) also needs to be taken into account. This means that both, the provider and the customer, intend to maximize their utility. In the following, we draw parallels between the decision factors identified and notions of RCT.

We have identified four groups of factors: service characteristics, provider interests, customer interests, and market factors. The revenue mechanism is chosen for a specific data-driven service, which is driven by certain characteristics. We can see that these specific characteristics itself have an influence on the choice of the revenue mechanism. Further, both provider and customer have individual interests and preferences regarding the revenue mechanisms that influence the choice. Finally, provider and customer interact in the context of an industry or market, which may have an influence on the selection of the revenue mechanism as well.

4.1 Service Characteristics

Data-driven services are such that rely on data as their key resource and analytics to provide value to customers. The offered services, however, may substantially differ from each other: for instance, there are heavy equipment manufacturers that start providing monthly usage reports to the operator of the machines for predictive maintenance purposes or a mobile phone network operator that provides targeted advertising services based on customer movement data. The nature of these offerings in itself can differ substantially and have an influence on the selection of the revenue mechanism. We identified two characteristics that play a role in the decision-making process: the usage pattern and the level of integration with a core product or service.

The usage pattern describes the frequency the customer actually uses the offering, which is often defined by the service itself. For instance, an alarm service of machinery that has to continuously monitor and process data in order to detect abnormalities. Hence, the data-driven service is provided continuously and adds value not just at a certain point in time. The choice of the revenue mechanism should therefore reflect that value for the customer is created on a continuous basis. The product manager in manufacturing company IV emphasizes this point using a different example: "(…) for a dashboard service, a pay-per-use model does not make sense since looking at the dashboard once represents a use. But this is not how a dashboard works. Dashboards show changes over time and need to analyze data continuously." In scenarios where the value of a service is usually derived from using it occasionally, e.g. generating a quarterly report on energy consumption from a utility provider, other revenue mechanisms may be more appropriate to use (e.g. usage-based).

Further, the integration with the core offering influences the choice of the revenue mechanism; i.e. the extent to which the data-driven feature is integrated with a core product or service. Data-driven offerings may be provided as a stand-alone service such as a navigation app on the smartphone or integrated with a core product or service such as system status monitoring for an elevator. A high level of integration between the data-driven service and the core offering makes it often difficult to distinguish the value created through the data-driven service from the one created through the core product or service. In such cases, services are often not charged separately and more likely to be charged indirectly through the revenue stream of the core product or service. With a decrease in the level of integration between the core offering and the data-driven service, there is an increase in the likelihood that an additional and therefore separate revenue mechanism for the data-driven service is chosen by the provider.

4.2 Provider and Customer Interests

Undisputedly, the provider of a product or service has an interest in capturing the value created by their offerings - not just to cover the costs but to create a sustainable business model with attractive profits. On the other side, customers that benefit from the offering and see how value is created for them are willing to pay for it. Therefore, the interests and preferences of the provider and the customer have an influence on the selection of the revenue mechanism. There are four dimensions – economical, relationship, capability, and common practices – within this group of factors that influence the choice of the preferred revenue. Economic objectives describe how financial targets that the provider and the customer deem relevant to their business strategy can influence the selection of a revenue mechanism. The relationship perspective between the two parties focuses on the level of trust and therefore supports or inhibits the implementation of certain mechanisms. Based on technical and knowledge capabilities, particular revenue mechanisms are enabled or prevented. Further, common practices outline habits and preference of the provider and customer for a revenue mechanism design.

Economical. While overarching financial objectives of organizations include the maximization of profits or revenues, operational targets can influence the selection of the revenue mechanism for data-driven services. Depending on how the management team sets these targets, certain models are more advantageous to implement than others. A project manager from an IT service company stresses the challenge in this: “Yes, we have also thought about it [a usage fee model], but we noticed that [with this model] we would place ourselves in a less favorable position (…) because of the usage behavior (…).” Financial objectives among our company sample vary broadly - even within one industry. Some organizations have a strong focus on ensuring that they have a quick and reliable return on investment since the setup of data-driven services often requires substantial upfront investments. For example, a product manager in manufacturing company II notes: “(…) I have to think from the perspective of a supplier. I must get a return on my investment. And for that reason, it is not important for me, if he [the customer] looks at the dashboard one time or one hundred times.” Consequently, the provider choses a mechanism where it can achieve a short-term return on investment. Furthermore, the ability of organizations to plan future cash flows with certainty is significant; organizations therefore oftentimes prefer mechanisms that have a fixed payment schedule (i.e. subscription model): “(…) a recurring payment is always attractive, simply because we have more predictable revenues, this holds true for the customer as well as for you, you have a continuous revenue stream, too. This way, both sides, provider and also user can plan the whole thing in a better way.” (Business Development Manager – Telco Industry). Providers need to ensure that running costs (e.g. server infrastructure) are covered; this is especially relevant for data-driven services that require to be available around the clock. Examples include the availability of a dashboard or an alert service. Subscription models are an example of a revenue mechanism that allows the provider to cover these running costs with a high level of certainty while e.g. usage-based models could lead to a gap in the revenue stream since they are less predictable.

Customers prone to risk avoidance may have an increased need for spending money in a very conscious way; i.e. customers only want to pay for the service when it is needed. A usage-based model is an example of a mechanism that supports this objective. Consequently, we can infer that the level of risk averseness of the provider and the customer alike define a personal preference in the decision process and therefore impact the order of preferred choices.

Relationship. The way the provider and customer regard and behave towards each other defines their business relationship. It is in the interest of the provider to build a long-lasting relationship in order to maximize the customer's lifetime value. The level of trust between the business partners is one factor that influences the strength of the relationship. Building trust into the provider to deliver the service in quality comes over time and sometimes requires the provider to give away a service for free at first before being able to charge for it. A product manager in manufacturing company I notes: "(…) we have done it in a way where, after installing the heavy equipment machinery, we offered it [the service] for free for two years and then started charging for it". The quality of the relationship, hence, enables or inhibits the use of certain revenue mechanisms. For example, a performance-based revenue mechanism can only be applied if there is a high level of trust: "(…) as long as I cannot prevent, on a technical level, that nobody can manipulate [results], in their favor, then such models [performance-based] are only possible with mutual trust" (Product Manager – Manufacturing Company VI).

The examples show that the state of the relationship between the partners has an impact on the choice of the revenue mechanism and therefore acts as a constraint in the process. Providers can treat customers fairly and build a trusted relationship to ensure ongoing revenue streams or follow a strategy to extract the maximum amount of revenue and profit from the client while accepting that it may not return for re-purchase.

Capabilities. The complexity of a revenue mechanism and its initial definition, implementation and monitoring can vary broadly. Simple mechanisms, such as a subscription, are often better received and understood by the customer compared to more complex constructs (e.g. performance-based). Furthermore, a lack of availability of knowledge and tools to implement the more complex models further limits the selection of revenue mechanisms available. Therefore, the existing capabilities on provider as well as customer side enable, limit or restrict the implementation of revenue mechanisms. The head of technology of an IT services company describes these limiting constraints in the choice process: "We didn't have an advanced or automated billing system to do usage-based or performance-based billing, so we just charged customers based on a simple monthly subscription in the contract agreement – which was simple for everyone to understand, and also simple to implement."

Data-driven services offered by the provider support one or multiple business processes on the client side. In order to setup more complex revenue mechanisms (e.g. performance-based), the client has to have a good understanding of how the service interacts with its processes and impacts business outcomes. If this is not given, more simple revenue mechanisms should be applied (e.g. subscription model).

Common Practice. We identified three levels where habits and common practices influence the choice of the revenue mechanism: individual, organizational and industry. On an individual level, we can observe that personal preferences of the person in charge of making strategic revenue model decisions plays a role. The person may transfer personal experience into the business environment and require a service to be offered in conjunction with a particular revenue mechanism.

Within an organization, there is a tendency to apply the same revenue mechanism for new data-driven services as has been done for existing ones. On the one hand, capabilities and processes for the implementation of existing revenue mechanisms are likely to be already in place and, therefore, a smooth implementation can be ensured. The introduction of a new model, on the other hand, oftentimes requires the definition and implementation of new processes, which may create additional risk of failure both for the provider and the customer. Hence, similar to the economical perspective, the level of risk avoidance and therefore a personal preference is a driving factor for the choice of a revenue mechanism.

On an industry level, customers expect the availability of certain mechanisms as a consequence of being common practices. For example, customers may expect a service to be offered for free when an expensive piece of heavy equipment machinery is purchased. An interviewee from manufacturing company III explains: "(…) it's sometimes difficult to point out to the customer that he should pay so much for this [the service], because says, 'if I buy such an expensive machine from you, then it [the service] should be included'" (Head of Innovation/Strategy - Manufacturing). This describes a scenario where the service is bundled with a core product – described as "wrapping" by Wixom & Ross [4].

The shift of common practices and therefore the preference of customers within an industry can also require the provider to introduce new and unproven revenue mechanisms. For example, the customers' intent to shift the risk of service fulfilment and success towards the provider. An example of a revenue mechanism that helps achieve this objective is the performance-based one since the provider only gets paid in case of proven success. Failure of a provider to offer a revenue mechanism that enables the shift of risk towards the provider side may inhibit the sale of the service.

Despite the fact that the provider is the one to make the final decision on the revenue mechanism design, customer preferences need to be considered by the provider. Failure to do so may result in customers choosing the offering of a competitor or not making a purchase decision at all, which, in turn, impacts the net benefit of the provider.

4.3 Market Factors

While provider and customer are most directly involved in the value creation and value capture of the data-driven service, they do not interact in a vacuum. There can be additional players involved that have an influence on the selection of the revenue mechanism. The behavior of competitors can urge the provider to offer one revenue mechanism over another. Further, the collaboration with partners in an ecosystem may require the provider to align its revenue model design with that of other players.

Therefore, market factors are to be considered a constraint in the decision of the agent and, according to RCT, limit the number of choices available to maximize utility.

Competitors. While it is often difficult to compare data-driven services between providers due to their unique and new character, bidding situations sometimes allow to observe what competitors offer. In addition, some revenue mechanisms may be more frequently used than others. An interviewee describes this as: "This is common in this industry" (Head of Innovation/Strategy – Manufacturing Company III). Being able to offer a revenue mechanism that is not common (e.g. performance-based) in a particular setting or industry may be recognized as a competitive advantage since special capabilities are often needed for the implementation.

Partners. Oftentimes, providers of data-driven services do not have the internal capabilities to develop and run a service on their own (e.g. hosting & connectivity services) and hence are required to collaborate with sub-providers. Therefore, the provider is urged to pick a revenue mechanism that ensures a continuous cash flow (e.g. subscription model) to cover ongoing obligations towards its own partners. The product manager of manufacturing company I describes this using an analogy: "(…) like the landlord of a building has to decide which costs to absorb and which to pass on to the tenants [to remain solvent]."

Many services nowadays are sold through 3rd party platforms (e.g. Apple Store, Google Play Store). These platforms may constrain the revenue mechanisms that are allowed to be used through contracts with the provider of the data-driven service. In addition, offering bundles offerings with a sales partner may further constrain the choices since the strategy needs to be aligned with other parties' expectations.

5 Conclusion

In summary, our research explored the factors that influence the selection of a revenue mechanism for data-driven B2B services. In order to do so, we conducted 14 expert interviews and analyzed them by applying open and axial coding. This led to the identification of four influencing factors: service characteristics, provider interests, customer interests, and market factors. Provider and customer interests as well as market factors are further broken down into subcategories to account for the specifics of each influencing criterion.

Each of these factors influence the provider when designing a revenue mechanism. The particular shaping of some factors even enable, hinder or promote certain revenue mechanisms (e.g. the lack of technical capabilities or knowledge to define a unit of measure inhibits the implementation of usage- or performance-based models). The choice of a revenue mechanism requires the provider to synthesize all available information and to decide on the best strategy for the situation at hand.

Our contribution with this paper is two-fold: on the one hand, we derive factors that influence the decision-making process for revenue mechanisms for data-driven services, and, on the other, we point out a set of initial schemes on the direction that each factor may influence a decision for or against a particular mechanism. With this, we contribute to the general understanding of revenue models and lay the foundation for more elaborate research in the field that could eventually help identifying the appropriate revenue model for a data-driven service.

5.1 Discussion

Looking at the influencing factors derived from the interviews through the lens of rational choice theory shows that a major part of the results (provider and customer interests, market factors) can be explained by either personal preference of the decision-making unit – informed by their attitude towards risk - or a constraint. However, not all results can be linked to RCT: service characteristics shape a category of their own and therefore stand out from the rest. These insights extend the knowledge in the field of revenue model research for data-driven services and contribute to developing a better understanding of decision-making for such services. It shows that the characteristics of the service being offered has a particular impact on the revenue mechanism. Thus, there is an opportunity to identifying the most appropriate revenue model for a particular service. Our study extends extant literature in the field of data-driven service research and provides the basis for more research in this space.

5.2 Managerial Implications

The implications of this study for practice are relevant in such that the design of the value capturing mechanism, which the revenue mechanism is a part of, has an influence on whether an offering becomes a success for the organization or not. Our results show that the selection of the revenue mechanism is a complex and critical endeavor due to the number and variety of factors that need to be considered. We have shown that some of the factors not only call for the use of a particular revenue mechanism, but that they can also preclude their use. Organizations should be aware that not only their own interests, but an entire ecosystem of influencing factors play a role when designing an overarching revenue model for data-driven services.

5.3 Limitations and Future Research

Despite applying a high level of rigor, our analysis is not without limitations. The services analyzed in the context of this study all apply to business-to-business transactions. We encourage to repeat this study taking also B2C services into account and to compare the results. Furthermore, our sample was limited to a total of four industries. Extending the study to additional industries would allow verifying the results in a broader context.

We encourage future research on the topic of revenue mechanisms for data-driven services. From our point of view, there are several areas that would benefit from further research. For once, our study focused on the identification of influencing factors for the selection of revenue mechanisms, however, future research could further explore the importance of each of these factors; i.e. to conduct a quantitative study on the subject.

A further avenue to explore is to develop an understanding if the factors identified also hold true in the context of B2C offerings since we focused B2B offerings in this study.

References

1. Brown, B., Chui, M., Manyika, J.: Are you ready for the era of 'big data.' McKinsey Q. 4, 24–35 (2011).
2. LaValle, S., Lesser, E., Shockley, R., Hopkins, M., Kruschwitz, N.: Big Data, Analytics and the Path From Insights to Value. MIT Sloan Manag. Rev. 52, 21–31 (2011).
3. Davenport, T.H.: Analytics 3.0. Harv. Bus. Rev. 91, 64 (2013).
4. Wixom, B.H., Ross, J.W.: How to monetize your data? MIT Sloan Manag. Rev. 58, (2017).
5. Hartmann, P.M., Zaki, M., Feldmann, N., Neely, A.: Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. Int. J. Oper. Prod. Manag. 36, 1382–1406 (2016).
6. Neely, A.: Exploring the financial consequences of the servitization of manufacturing. Oper. Manag. Res. 1, 103–118 (2008).
7. Chesbrough, H., Rosenbloom, R.S.: The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. Ind. Corp. Chang. 11, 529–555 (2002).
8. Teece, D.J.: Business Models, Business Strategy and Innovation. Long Range Plann. 43, 172–194 (2010).
9. Johnson, M.W., Christensen, C.M., Kagermann, H.: Reinventing Your Business Model. Harv. Bus. Rev. December, 1–10 (2008).
10. Schüritz, R., Seebacher, S., Satzger, G., Schwarz, L.: Datatization as the Next Frontier of Servitization – Challenges of Organizational Transformation. In: Proceedings of 38th International Conference on Information Systems (2017).
11. Ulaga, W., Reinartz, W.J.: Hybrid Offerings: How Manufacturing Firms Combine Goods and Services Successfully. J. Mark. 75, 5–23 (2011).
12. Schüritz, R., Seebacher, S., Dorner, R.: Capturing Value from Data: Revenue Models for Data-Driven Services. In: Proceedings of 50th Hawaii International Conference on System Sciences. pp. 5348–5357 (2017).
13. Sprenger, M., Mettler, T., Winter, R.: A viability theory for digital businesses: Exploring the evolutionary changes of revenue mechanisms to support managerial decisions. Inf. Syst. Front. 19, 899–923 (2017).
14. Pantelis, K., Aija, L.: Understanding the value of (big) data. In: Proceedings of IEEE International Conference on Big Data. pp. 38–42 (2013).
15. Chen, H., Chiang, R.H.L., Storey, V.C.: Business Intelligence and Analytics: From Big Data To Big Impact. MIS Q. 36, 1165–1188 (2012).
16. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity. (2011).
17. Yang, Z., Djurdjanovic, D., Ni, J.: Maintenance scheduling in manufacturing systems based on predicted machine degradation. J. Intell. Manuf. 19, 87–98 (2008).
18. Osterwalder, A.: The Business Model Ontology - A Proposition in a Design Science Approach, (2004).
19. Barquet, A.P.B., de Oliveira, M.G., Amigo, C.R., Cunha, V.P., Rozenfeld, H.: Employing the business model concept to support the adoption of product-service systems (PSS). Ind. Mark. Manag. 42, 693–704 (2013).

20. Sawhney, M.: Going Beyond the Product: Defining, Designing and Delivering Customer Solutions. In: The Service-dominant Logic of Marketing. pp. 365–380. M.E. Sharpe (2006).
21. Lay, G., Schroeter, M., Biege, S.: Service-based business concepts: A typology for business-to-business markets. Eur. Manag. J. 27, 442–455 (2009).
22. Al-Debei, M.M., Avison, D.: Developing a unified framework of the business model concept. Eur. J. Inf. Syst. 19, 359–376 (2010).
23. Shafer, S.M., Smith, H.J., Linder, J.C.: The power of business models. Bus. Horiz. 48, 199–207 (2005).
24. Kindström, D., Kowalkowski, C.: Service innovation in product-centric firms: a multidimensional business model perspective. J. Bus. Ind. Mark. 29, 96–111 (2014).
25. Van Ostaeyen, J., Van Horenbeek, A., Pintelon, L., Duflou, J.R.: A refined typology of product-service systems based on functional hierarchy modeling. J. Clean. Prod. 51, 261–276 (2013).
26. Bonnemeier, S., Burianek, F., Reichwald, R.: Revenue models for integrated customer solutions: Concept and organizational implementation. J. Revenue Pricing Manag. 9, 228–238 (2010).
27. Kim, S.-H., Cohen, M.A., Netessine, S.: Performance Contracting in After-Sales Service Supply Chains. Manage. Sci. 53, 1843–1858 (2007).
28. Ng, I.C.L., Maull, R., Yip, N.: Outcome-based contracts as a driver for systems thinking and service-dominant logic in service science: Evidence from the defence industry. Eur. Manag. J. 27, 377–387 (2009).
29. Keskin, T., Taskin, N.: A Pricing Model for Cloud Computing Service. In: Proceedings of 47th Hawaii International Conference on System Sciences. pp. 699–707 (2014).
30. Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D.G., Lewis, R., Rao, A., Sahni, N., Yao, S.: How do firms make money selling digital goods online? Mark. Lett. 25, 331–341 (2014).
31. Levin, J., Milgrom, P.: Introduction to Choice Theory. (2004).
32. Wiesche, M., Jurisch, M., Yetton, P.W., Krcmar, H.: Grounded Theory Methodology in Information Systems Research. MIS Q. 41, 685–701 (2017).
33. Krippendorff, K.: Content Analysis: An Introduction to its Methodology. Sage Publications, Thousand Oaks, California (2004).
34. Bengtsson, M.: How to plan and perform a qualitative study using content analysis. NursingPlus Open. 2, 8–14 (2016).
35. Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K.: Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm. Policy Ment. Heal. Res. 42, 533–544 (2015).
36. Introna, L.D., Wood, D.: Picturing algorithmic surveillance: The politics of facial recognition systems. Surveill. Soc. 2, 177–198 (2004).
37. Saldaña, J.: The Coding Manual for Qualitative Researchers. SAGE, London (2009).
38. Glaser, B.: Theoretical sensitivity: Advances in the methodology of grounded theory. Sociology Press (1978).

Understanding the Role of Data for Innovating Business Models: A System Dynamics Perspective

Matthias Förster[1], Bastian Bansemir[2], Angela Roth[1]

[1] Friedrich-Alexander University Erlangen-Nürnberg, Chair of Information Systems - Innovation & Value Creation, Nürnberg, Germany
{matthias.f.foerster,angela.roth}@fau.de
[2] BMW Group, Business Model Development, München, Germany
{bastian.bansemir}@bmw.de

Abstract. Data have become a key ingredient for ICT-enabled business models. Nevertheless, there is great uncertainty among scholars and practitioners alike about how to leverage data as an essential innovation resource. This raises the question of how to design data roles to innovate business models. To answer this question and to facilitate a deeper, cause-effect-relation understanding of the interdependencies between data roles and business models, system dynamics was chosen as the approach of analysis. Within a multiple case study of five business model cases with multiple embedded units per case, the study shows that there are four recurring basic data role patterns e.g. 'incremental improvement' or 'initial data boost' and two data role characteristic patterns, describing how data roles unfold within business models e.g. 'change in self-reinforcement' pattern. Overall, the patterns help to visualize and articulate data usage in business models and therefore contribute to the ongoing endeavor of innovating business models.

Keywords: Business Model Innovation, Big Data, System Dynamics, Data Role Patterns

1 Introduction

In the course of increasing digitalization, information and communication technology (ICT) has become both an enabler and constraint for new business models (BMs) [1]. ICT-enabled and dynamic business processes require a constant adapting and reshaping of BMs to cope with the continuously changing business environment across all industries [2, 3]. As a key ingredient or even ignitor for ICT-enabled innovating of BMs, 'digital data' have increasingly gained importance [4, 5]. This paper aims at facilitating a deeper understanding of how data innovate BMs, and shows the impacts of data on BMs. For this purpose, the paper examines the design of data roles, which are resulting from the usage of data to achieve an overarching business goal, e.g. efficiency gains through process transparency using IT systems, in order to innovate BMs. The term 'data roles' describes on a cause-effect-relation basis the task and underlying concrete impact of data on the elements of a business model (BM). Going one step further, the study enriches information systems (IS) research, with outlining

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

the reciprocal relationship between IT-capabilities, the choice of data roles, and the interlink to BM configurations. Thus, the paper positions itself at the vertex of IS and business model innovation (BMI) research, because so far literature largely and particularly discusses how to make data consistent with business strategy [6], reflects the opportunities of leveraging data to innovate existing BMs, outlines how to build completely new BMs [7] or internet-based BMs [8], and addresses constructing and implementing data related adaptions and amplifications to existing BMs [4, 9, 10]. However, in order to be able to build completely new 'data-based' BMs [11], an awareness on how to design data roles to innovate BMs with consideration of data impacts on the BM is required. The failure of building these completely new BMs may be ascribed to a lack of knowledge regarding BMs and the concatenating management of digital data as an essential resource to innovate BMs. This is not a surprise due to the specific and partly fuzzy characteristics of data which make data as a resource difficult to manage [12, 13]. Because of these complex data characteristics, it is not sufficient to give way to simplifications and generalizations [14, 15] at the endeavor to understand the role of data to innovate BMs. Moreover, to grasp non-linear, cause-effect-relations between data roles and BMs, a novel data-focused perspective on BMs is required. To meet these requirements and to provide managers and scholars an advanced understanding of the dynamic data roles behind the static BM story, illustrated as target images in a BM (data) canvas [16–18], a dynamic and conceptual representation method is premised. System dynamics - a computer-aided approach to enhance analysis and decision making for complex systems [19] - turns out to be a suitable approach for this, due to it's potential of simplified and consistent representations of BMs, including quantified considerations of feedback loops and delays between the individual BM elements [20]. In case of our study, the interrelations and feedbacks between data roles and BMs are focused. With use of the system dynamics approach the study seeks to answer the following research question: ***How could data roles be designed to innovate business models?***

2 Theoretical Background

2.1 Innovating Business Models

According to Shafer et al. (2005) [21] the study's central subject of matter – BMs – are defined as 'the firm's logic of creating value'. Although many authors conceptualized a holistic view on the logic of a firm's value creation by outlining consolidated components of a BM, a common understanding is not overarchingly achieved [22]. Examining the most cited publications on this topic one may conclude that there is common understanding among the following three components within the BM definition [11]: value proposition, value creation and value capturing [10, 23–28]. Gassmann et al. (2014) [17] breaks these three components down to four fundamental BM questions: "Who is my target customer?", "What is offered to the target customer?", "How to build and distribute the value proposition?" and "Why is the BM financially viable; what value is generated?" As the hackathon artefacts A1 and A4 of the empirical field (see figure 1) are based on the BM definition of Gassmann et al.

(2014), the BM understanding of this study is oriented on the conceptualization of the BM components: How (value creation), what (value proposition), value (value capturing) and who (key customer). Since the emergence of the internet economy the topic of innovating BMs has gained much popularity [29], and BMI can be seen as a subordinate BM research stream. With the opportunities of the internet and the accompanying ICT, firms obtained a broad armamentarium to completely change the rules of competitive environment [30] by approaching the customers with entirely new, data-based BMs [31]. However, innovating BMs remains an ambiguous concept. Thus, different research streams consider BMI either as a process (e.g., search, experiment, transformation), while another stream see BMI as a output. [32]. As our research is based on a hackathon where 171 students innovated BMs by initiating and ideating new BMs based on the firm profiles, capabilities and existing BMs, we understand BMI as a process. This is in line with the research question *"How could data roles be designed to innovate business models?"*, which also provides support for the ideation process of innovating BMs. Furthermore, system dynamics as a dynamic BM experimentation, modelling and simulation approach also understands innovating BMs as a process. In comparison, most conventional approaches fall short grasping the dynamics of BMs over time. They depict innovation as a static image or a 'snapshot' of a BM at a certain point in time, i.e. when it functions as wished. But a BM - especially a data-based BM [2] - is by no means a purely static concept. It must constantly be innovated and adapted to changing internal and external forces [25] in order to protect the recently implemented BM against imitations [10, 28], or react [32] to strategic discontinuities or global competition [33]. However, especially long-established firms, struggle with constantly innovating BMs due to the complexity of inherent IT-systems and - infrastructure or path dependency such as legacy systems and products [10]. Therefore, innovating BMs is often lengthy, risky or costly, or at worst, all three of the before mentioned challenges [34]. These challenges can be overcome and BMs can be made manageable by adopting an effectual attitude toward BM modelling and experimentation [35]. Chesbrough (2010) states: *"[Model] experiments will fail, but [if] they inform new approaches and understanding, this is to be expected - even encouraged"* (Chesbrough, 2010 p. 362). Also, Osterwalder et al. (2005) [36] claim that *"simulating and testing business models is a manager's dream"* (Osterwalder, 2005 p. 16). In other words, to overcome the above mentioned 'deadlock' for established companies and to keep pace with the dynamically changing ICT-enabled business environment, new data-related approaches for innovating BMs must be developed. BM simulation, more precisely the system dynamics approach, appears to be a suitable and necessary means to innovate BMs.

2.2 System Dynamics

System dynamics is a simulation and experimentation approach for enhancing analysis and decision-making on complex and dynamic systems [20]. System dynamics was initially developed by J. W. Forrester in 1968 [37] as a methodology developed from several scientific approaches including system theory, information science and cybernetics [38]. System dynamics is considering all elements of a system as actively influencing the behaviour of the system through interdependencies and mutual interactions of elements, so-called feedbacks. The cyclic consequences inside a system are represented by causal-loop diagrams (CLDs). A CLD shows a sequence of events (actions, information, objects, people) successively causing another event until the first mentioned event is caused again. These feedback loops can either be self-reinforcing and dominating the system over time due to exponential growth effects, or self-correcting and balancing a system due to corrective and attenuating effects [39]. The CLD visualization of several reinforcing or balancing, non-linear, multi loop conditions help people to remove cognitive barriers for understanding the complexity of a system. For managing highly interrelated, complex systems such as BMs, system dynamics uses computer-aided modelling and simulation tools like Powersim or Vensim. These computational simulation tools help managers to quantify flows (information, goods, people or financial means) inside BMs, and therefore, provide managers and scholars the opportunity to test and gauge their decisions and to learn about the consequences of their actions in complex corporate and industry contexts of their BMs [20, 37]. Understanding, testing and gauging BM decisions is especially important during the ideation phase of innovating BMs where managers generate new BM ideas by overcoming the current BM logic through reflecting novel, non-trivial changes in the way the BM elements are interrelated [32]. System dynamics has been applied to various BM domains like transportation [38], economic and ecological sustainability [40] or the revolutionary and evolutionary market strategies at the automotive security equipment sector [41].

3 Research Design

3.1 Case Study Design

To explore how data roles could be designed to innovate BMs an exploratory multiple case study was carried out. The case study method aims at retrospectively comprehend complex issues of a contemporary phenomenon within a real-life setting. Furthermore, the case study method is useful when literature and theoretical knowledge do not fully delineate the issue of research with appropriate certainty and clearness [42]. We see these aspects of a multiple case study as given in our study, since the case units of the case study were carried out in the context of an industry-initiated and -derived student hackathon with real-life oriented design decisions in the hackathon process to increase the significance for industry. Additionally, the process was accompanied and supervised by a total of three industry partners. The empirical field of the multiple case study was a hackathon, a collaborative event where 33 student teams ideated BMs for the mobility sector. These 33 teams were split to five different cases, which represented the five essential actors in the mobility sector. Each actor had different pre-settings through company history, firm resources and capabilities, existing BMs and financial KPIs. These actors were: an automotive OEM, mobility fleet operator, digital mobility aggregator, public transportation company and a new automotive rebel OEM. These five mobility actors were represented multiple times as embedded units per case.

3.2 Hackathon Design

Hackathon Approach

A hackathon is a marathon coding event in which a group of predominately computer programmers or IT-designers are intensively involved in developing a software prototype for a specific application, or applying a technology prototype for a specific purpose over a short period of time [43–45]. In this study, software developments or technology applications (the pure coding) were in the background. The objective of the hackathon was to let student teams compete against each other in a partially defined competitive environment with the task to come up with innovative data-based BMs Hence, to quickly and interactively generate and test BMs for real-life problems in a simulation environment [44, 45] - in our case the manifestation of future mobility.

A hackathon can be carried out with different target groups according to the statement of problem and the desired prototype output [46]. For this hackathon, student teams represent a scientifically significant and suitable hackathon execution group for two main reasons: Firstly, they are unbiased and open minded regarding new, disruptive ideas because of no long-term organizational affiliation. Therefore, more distinguishable results can be expected than with a purely industry hackathon. Secondly, to meet the needs of future customers by means of new innovative BMs, it is essential to include today's students in the innovation process of these BMs.

Future Mobility Hackathon

The Future Mobility Hackathon was a collaborative project between two university institutions of Switzerland and a German car manufacturer to foster creativity and a spirit of innovation among students while leveraging the power of system dynamics, and to support the practice partner to gain an advanced understand of future mobility needs. A total of 171 participants from both university institutions worked in teams to ideate, model and analyse innovative but also viable BMs for the mobility sector. The hackathon was conducted on a total of four dates during October and November of the year 2017. Throughout the hackathon the teams developed a series of six BM artefacts to go from an innovative idea and target image of the BM, based on the BM definition concept of Gassmann et al. (2014) [17] (BM magic triangle) to graphical illustrations and evaluation models for the dynamics of the BM (CLDs). The hackathon concept was fitted by industry-derived design decisions like different pre-settings of the hackathon players e.g. firm history, resources and capabilities, and university program-related design decisions like the release of the hackathon competition evaluation criteria at the beginning of the event to ensure the industry significance and to support the students in two ways: reducing the complexity of innovating BMs, and increase the students' identity with the object of study and therefore their creativity in the innovation process.

3.3 Data Selection and Collection

As stated above, five different hackathon BM cases, each with multiple embedded units (in total 33 teams) were examined in the multiple case study according to the case study method of Yin (2012) [42], which can be divided into 3 main phases: 1) define and design, 2) prepare, collect and analyse, and 3) analyse and conclude. In the first phase the data selection was carried out. Impartially of the team performance in the hackathon, the authors selected all generated artefacts (6 artefacts per team; 33 teams in total) throughout the hackathon process. The data collection (case study method phase 2) took place based on several data sources (see figure 1): sketched and computer-aided artefacts (A1-A6) ranging from a rough description of the target image of the BM by means of the BM magic triangle [17], over differently sophisticated CLDs of the BMs (which show the dynamics of the teams' ideated BMs) to the BM pitch presentations in the final competition round. The artefacts (A1-A6) were collected through submissions at the end of each hackathon competition phase. During the hackathon phases where the teams actively ideated, conceptualized and adjusted their artefacts, qualitative team observations were carried out by 6 hackathon supervisors. The focus of these team observations was on the team debates regarding data inside the artefacts and on the team's notion about how data roles could be designed to innovate BMs. These qualitative team observations were an important part of the data collection to give the authors a deeper understanding of data roles in BMs and to find out if there were conflicting views on the usage and role of data. Figure 1 explains for each phase of the hackathon the outcomes of the hackathon, which way they were collected (submission, team observations, or semi-structured interviews) and their input for the case study.

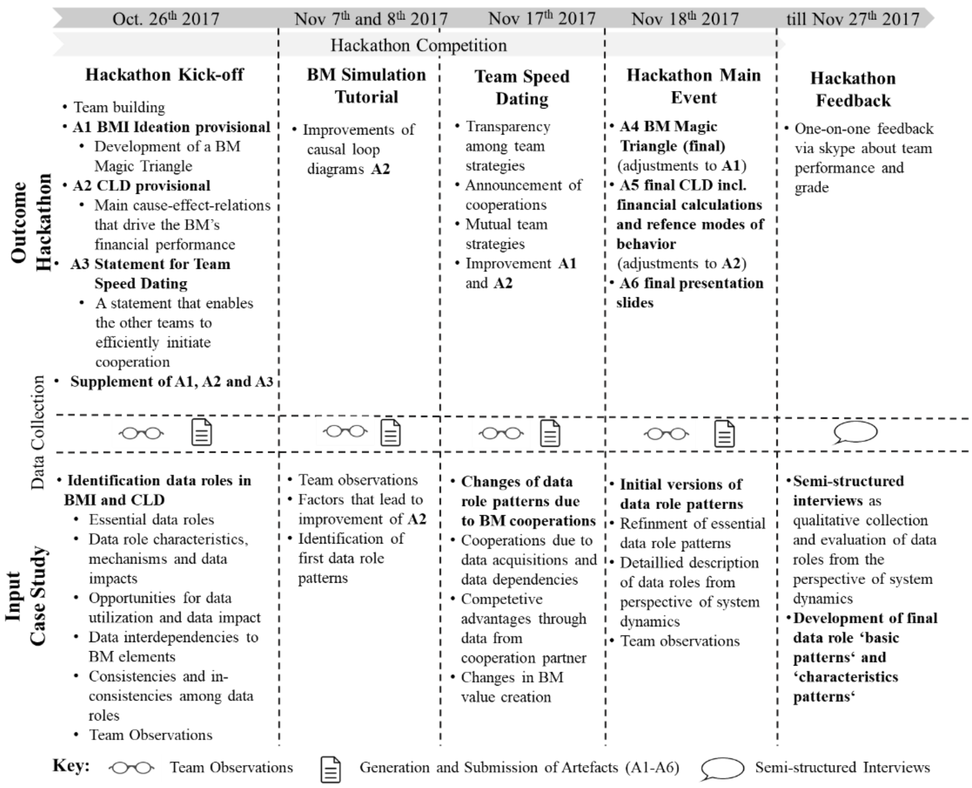


Figure 1. Data collection of the case study

Based on the findings from the analysis of the artefacts as well as the team observations, semi-structured interviews [47] were conducted with 24 of 33 hackathon teams in the aftermath of the hackathon. Since the interviews were not mandatory for the students, not all student teams agreed to be interviewed. The aim of the semi-structured interviews was firstly to uncover personal experiences, knowledge and attitudes of the hackathon participants regarding the design of data roles to innovate BMs. Secondly, the semi-structured interviews aimed at validating and enriching data from the preceding analyses of the hackathon artefacts and team observations to better explain data roles in BMs in order to obtain more objective findings as a valid foundation for further research. The 24 interviews were conducted based on a questionnaire, consisting of five open-ended questions. The questions were arranged in a way that at first the interviewees were free to state and explain the roles of data in their BM as they thought of them, how important they consider these data roles to innovate BMs, and on how they designed and realized them in their artefacts. These questions were followed by more specific questions in which the interviewees had to find and explain data roles, their characteristics and impacts on BM elements inside their CLD artefacts.

3.4 Data Analysis

We started our analysis for *the design of data roles to innovate BMs* with reflecting the BM magic triangle artefacts A1 and A4 in order to understand the BM target images and why data were used within BMs to get a first sentiment on data roles (see figure 2 step 1). To enhance the quality of this step of analysis and to check if there were conflicting views on the usage and role of data within the team, or if we might have misunderstood the artefacts, we cross-checked them with the qualitative team observations. In case that the teams have expanded their BMs by collaborating with other teams, we also analysed the artefacts A3, because collaborations may have an impact on the use of data. In step 2 we analysed the CLD artefacts A2 and A5 to understand the BM dynamics, reasons for data usage and positions of data as well as their impacts on the BMs. Moreover, we analysed the dynamics and consequences of data impacts (data loops) on BMs. In step 3 we extracted the identified data loops and cross-checked their validity, objectivity and reliability with the data loops inside and outside the cases. Again, we cross-checked the data loops with qualitative team observations and afterwards we validated and enriched our findings with data from semi-structured interviews. Based on step 1-3 we developed the six abstracted and generalized data role patterns that describe how data roles could be designed to innovate BMs (step 4). Figure 2 shows the 4-step analysis that identified the data role patterns.
Using the example of the illustrated 'the more the more' data role pattern at figure 2, we will briefly explain how to read our findings (data role patterns). The circle, called 'loop', represents the core mechanism of a BM. The general rule of system dynamics states that any intervention of an event triggers effects that eventually refer back and cause the same event again. This coherent relation is illustrated through circles (loops) based on several individual 'events'. The data-affected loops can be self-reinforcing, represented by a plus with a circular arrow around it or balancing, represented by a minus with a circular arrow around it. For matters of reading comprehension, the

authors recommend reading the abstracted and generalized data role patterns, in the direction of the data impact (data impact direction is highlighted by bold arrows). In case of the above stated loop this means: The more useable data (e.g. infrastructure, vehicle, road environment) is generated through a higher utilization rate of shared autonomous driving cars, the more [higher] data quality can be ensured. More [higher] quality improves the relative level of competition and thereby attracts more customers, which lead to a higher use of the service, and thus to a higher data generation, with which the quality of the shared autonomous driving service can be improved again. The loop is self-reinforcing.

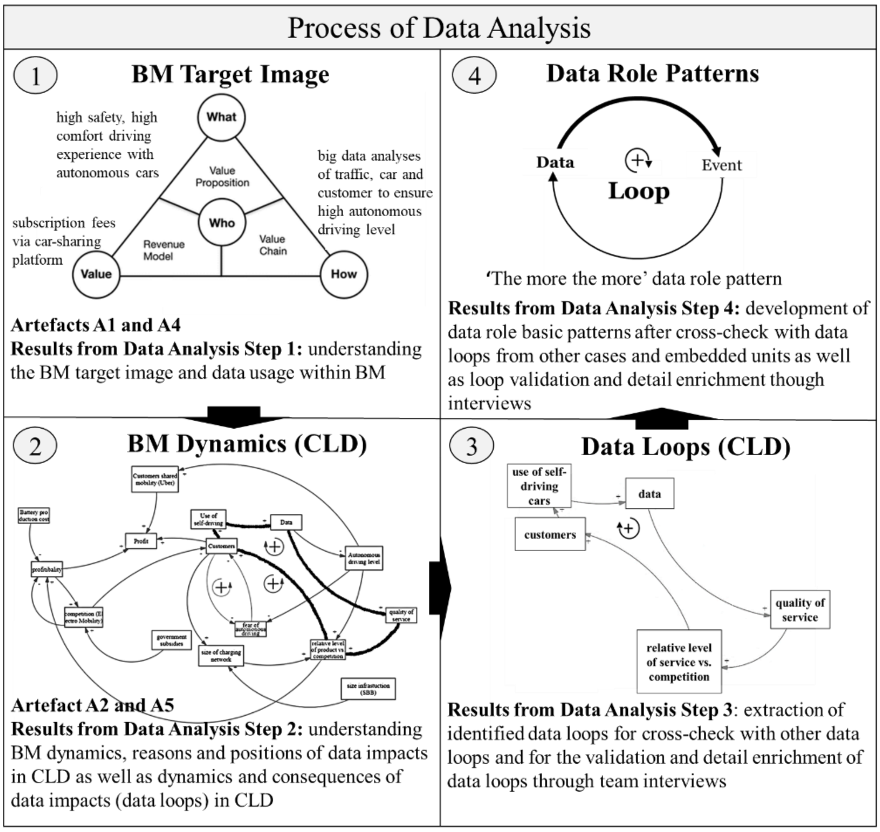


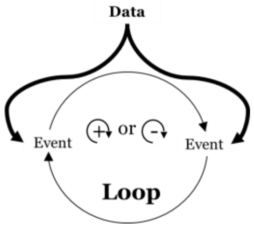
Figure 2. Process of data analysis

4 Findings

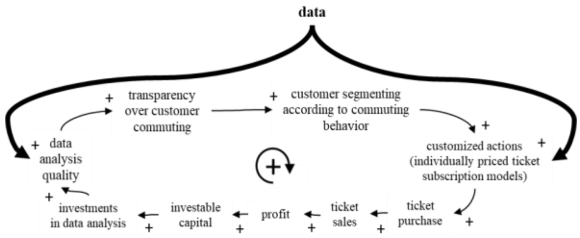
This chapter outlines the six key findings of the multiple case study. According to the research question "*How could data roles be designed to innovate business models?*" the authors identified the following recurring six data role patterns. These are: **'incremental improvement'** (**KF1**), **'initial data boost'** (**KF2**), **'business enabling'** (**KF3**) and **'the more the more'** (**KF4**). As all five cases of the case study with their multiple embedded units compose of these four data role patterns, they are called ‘data role basic patterns’. Moreover, the study identified certain, complex characteristics of these patterns, describing the combination of the four ‘data role basic patterns’ (**‘mix and match’** (**KFa**), and the dynamics and thus implicit change in the mode of action of the patterns (**‘change in self-reinforcement’** (**KFb**). These patterns are called ‘data role characteristic’ patterns. The following tables (table 1 & 2) represent the key findings of the study with brief descriptions of their data role mechanisms, revised snippets of the original CLDs and the corresponding abstracted and generalized graphical data role patterns.

Table 1: Data role basic patterns

Key Finding 1: ‘Incremental improvement’ data role pattern



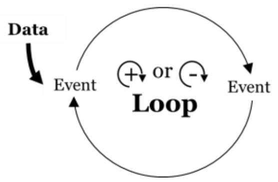
KF1 (#21)



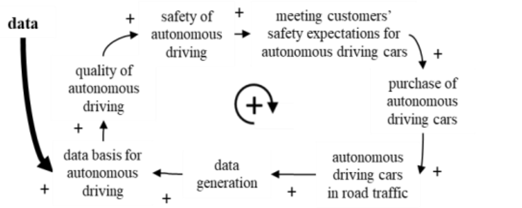
Description: Data are used to incrementally enhance the way a BM creates, captures or proposes value through data impact on one or more events of an existing BM loop. Data only induce incremental business improvements like enhanced transparency for customized actions. However, the traditional BM is only partially innovated.

Example: *Individually priced tickets: If a customer uses an app or the computer to book the daily subway ride from home to work, the transportation company knows through data analysis that the customer commutes daily, and thus can offer the customer an individually priced subscription. The BM is just incrementally improved and not fundamentally reinvented.*

Key Finding 2: ‘Initial data boost’ data role pattern



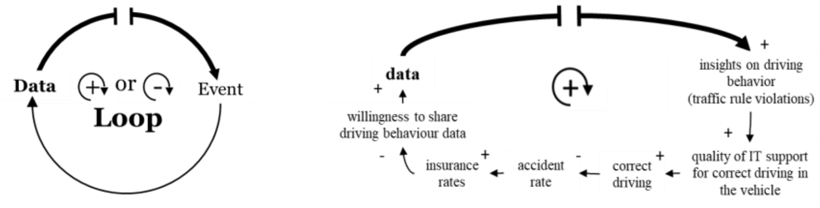
KF2 (#14)



Description: Through an initial, intense data-boost the events of the core mechanisms of a BM will be enriched with the necessary critical amount of data. Once the critical data amount is reached the functioning of the essential loops is initiated and the BM starts operating.

Example: *Safety of autonomous driving cars: If customers are willing to use autonomous cars in case that they are safer than non-autonomous cars, an initial high data volume (data boost) is necessary to ensure customers' minimum safety expectations of the service. Once the certain safety expectations have been met, the safety of the service doesn’t need further significant increases.*

Key Finding 3: 'Business enabling' data role pattern

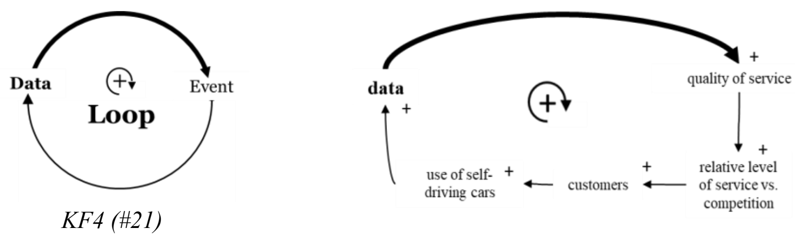


KF3 (#21)

Description: Data enact as business enabling because the BM is only made possible by the utilization of data, e.g. through data insights. Compared to the 'initial data boost' data role pattern, at this data role pattern, the core mechanisms of the BM have a constant, equally intense impact and dependency on data.

Example: *<u>Pay how you use insurance rates</u>: Instead of annual vehicle insurance rates, customers pay proportionally based on their driving behaviour. Traffic rule violations are displayed directly in the vehicle to educate the driver. Unsustainable and vehicle-damaging driving is penalized by higher rates; correct driving is rewarded with lower unsurance rate. The accident risks for the insurance companies decline. Thus, they can offer cheaper rates. Cheaper rates increase the customes' willingness to share driving behaviour data. The loop is self-reinforcing.*

Key Finding 4: 'The more the more' data role pattern



KF4 (#21)

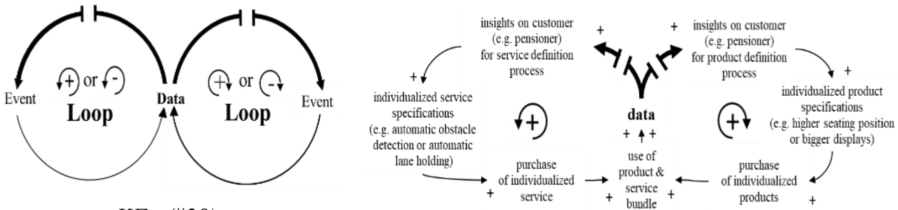
Description: BMs can have a 'the more the more' data role pattern e.g. regarding the quality of the offered service (value proposition) if 1) the quality of the service is directly improved by the quality of the database, and 2) the quality of database is improved by the amount of available data. 'The more the more' data role patterns are self-reinforcing.

Example: *<u>Service quality of autonomous driving</u>: The more an autonomous driving service is used, the more data are generated. With more data the algorithms for autonomous driving are getting better, and thus the quality of the service is getting better. The better the quality of the service, the higher the relative level of competition. Therefore, more customers use the service, and subsequently the loop reinforces again.*

As stated above, the study also identified two patterns which describe the characteristics of the four 'data role basic patterns'. These so-called 'data role characteristic patterns' (see table 2) are '**mix and match'** (**KFa)**, and '**change in self-reinforcement'** (**KFb**).

Table 2. Data role characteristic patterns

Key Finding 5: 'Mix and Match' data role characteristic pattern

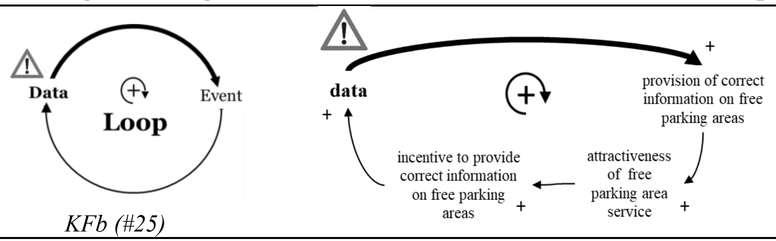


KFa (#28)

Description: On prominent characteristic of the four basic data role patterns is that they can be combined with each other in any form and number. In the exemplary graphical illustration two 'business enabling' data role patterns are arranged around data as the central point of both loops. Because of the centricity of data (two loops), data have an enormous impact on two complementary core mechanisms within BMs through simultaneous direct effects, and thus significantly determine the dynamics of a BM.

Example: *Individual cars with individual mobility services: Based on customer insights (e.g. customer is a pensioner) the firm can individually design cars according to customer preferences e.g. higher seating position and larger displays. Also based on customer data, the firm can offer complementary mobility services for the individually designed cars like lane holding assistants. Thus, the firm uses data to enable two different but complementary BMs, building an individual mobility environment for the customer.*

Key Finding 6: 'Change in self-reinforcement' data role characteristic pattern



KFb (#25)

Description: The pattern describes a possible dynamic behavior of a self-reinforcing data role over time. With changing data attributes as volume or quality, the role of data can change abruptly in the way it is operating in reinforcing loops. Therefore, a business-fostering, self-reinforcing loop changes to a business-destructing, self-reinforcing loop. Consequently, data become an exponentially growing threat to the BM. This pattern is an enormous uncertainty factor for BMs, because once an input flow (e.g. data volume) changes, the exponentially increasing consequences are hard to handle.

Example: *App-based free parking area community service: A service where members tell each other about free parking areas in the city, is business-fostering, self-reinforcing until false data (e.g. by opportunism of residents in free parking streets) is distributed to the community network. False information leads to a lower attractiveness of the service, which reduces incentives for other users to make their known free parking areas available to the community. As a result, there is less correct information provided in the app and the attractiveness of the BM continues to decline exponentially (business-destructing loop).*

5 Discussion and Conclusion

The study aimed at facilitating a deeper understanding of how to leverage data as a new essential resource to innovate BMs. For this purpose, the paper examined - on the basis of an exploratory multiple case study [42] - the design of data roles to innovate BMs. Through the grounding on system dynamics [20] the identified data role patterns (KF1-KF4, KFa & KFb) provide an advanced understanding of the complex and dynamic data impacts and concrete interrelations with BMs. Moreover, the identified six data role patterns take the discussion of *"how to leverage data to innovate BMs"* one step further and provide generalized and consistent representations of how to design data roles for the viability of BMs. This distinguishes the study from previous contributions, which rather reflect the nature of data usage and the potential and importance of data for the core logic of a particular BM from a more static and conceptual level [7, 9, 48]. However, some may argue, that the patterns are relatively established knowledge, debating network effects, platform BM interdependencies or critical data masses for market strategies, but the authors are convinced that the paper is of novelty for the discussion of how to innovate BMs with data for three reasons: 1) visualization of data impacts, 2) possibility of data impact quantification, and 3) validation of known data roles in data-based BMs outside the realm of purely digital BMs. In detail:

Firstly, the data role patterns help to visualize complex, cause-effect-relations to contribute to an enhanced management articulation of data usage. Especially consequences of data usage to improve value creation, value capturing and value proposition via building a clear step by step chain of effects is of particular interest for scholars and managers alike. Therefore, the patterns help to overcome cognitive barriers at decision making both in highly volatile industries like emerging ‘blue oceans’ [49] and existing industries in transition such as mobility sector, health care, insurance, aviation or mechanical engineering. Secondly, applying system dynamics approach, managers, consultants and researchers are able to quantify data role impacts to test and gauge data usage to design the data roles best suited for the desired BM. Thirdly, the study explores that known cause-effect-relations of data roles from the realm of purely digital BMs, also apply at data-based BM in the industrial sector, where customer needs (e.g. mobility) are still realised by physical (non-digital) products like vehicles or trains. Hence, the findings contribute to an advanced understanding and management of designing data roles to innovate BMs impartially, be it a long-established industrial firm (e.g. an OEM, which integrate digital services in their current product offerings and IT-capabilities), a start-up /digital rebel (which builds BMs from scratch due to the utilization of data) or a highly digitalized incumbent firm (which expands its existing digital portfolio in new domains).

Also, the study has implications for scholars, because it provides a new, conceptual view (data role patterns) on a largely discussed issue [9-14] namely, how to leverage data to innovate BMs. Thus, the study builds i.a. on the research of Casadesus-Masanell and Ricart (2011) [50] on how to design a ‘winning BM’. Exposing causal insights on how to design data roles to innovate BMs, this study is giving a knowledge foundation on answering future research questions on how to design a winning ‘data-based’ BM. In accordance with the design decisions of the hackathon, the findings of this study are bounded. Therefore, further research projects should validate the identified patterns and search for additions and modifications under different settings of the empirical field. Due to the methodological approach of a multiple case study [42], qualitative data is collected and analysed. Future studies should carry out quantitative analyses to explore and quantify the intensity of data role impacts on BMs. Also, this study contributes to the ongoing research by setting out a first explanatory approach of how to design data-based BMs, considering data as a dynamic BM resource that is causally intertwined with the BM components value creation, value proposition and value capturing [11].

Analysing the data role patterns in light of resource-based view [51], i.e. understanding data as an existing or evolving resource, or a dynamic-capability [52] i.e. the management of data roles over time, may add considerable theoretical and conceptual insights, leading to fascinating and inspiring research questions and empirical approaches.

References

1. Bouwman, H.: The sense and nonsense of Business Models. Position Paper (2002)
2. Al-Debei, M.M., Avison, D.: Developing a unified framework of the business model concept. European Journal of Information Systems 19, 359–376 (2010)
3. Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J.M., Loos, P., Spann, M.: Geschäftsmodelle. Wirtschaftsinformatik 56, 55–64 (2014)
4. Brownlow, J., Zaki, M., Neely, A., Urmetzer, F.: Data and Analytics. Data-Driven Business Models: A Blueprint for Innovation. Unpublished, Cambridge (2015)
5. Iansiti, M., Lakhani, K.R.: Digital Ubiquity: How Connections, Sensors, and Data Are Revolutionizing Business. Harvard Business Review 92, 90–99 (2014)
6. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.: Digital Business Strategy. Toward a next generation of insights. MIS Quarterly 37, 471–482 (2013)
7. Wamba, S.F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D.: How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics. 165, 234–246 (2015)
8. Tucci, C.L., Brews, Peter, J.: Building Internet Generation Companies: Lessons from the Front Lines of the Old Economy. Academy of Management Executive 17 (2003)
9. Hartmann, P.M., Zaki, M., Feldmann, N., Neely, A.: Capturing value from big data – a taxonomy of data-driven business models used by start-up firms. International Journal of Operations & Production Management 36, 1382–1406 (2016)
10. Johnson, M.W., Christensen, C.M., Kagermann, H.: Reinventing Your Business Model. Harvard Business Review 86, 50–60 (2008)
11. Schüritz, R., Satzger, G.: Patterns of Data-Infused Business Model Innovation. 18th IEEE Conference on Business Informatics 18 (2016)
12. Lycett, M.: 'Datafication'. Making sense of (big) data in a complex world. European Journal of Information Systems 22, 381–386 (2013)
13. Yoo, Y., Lyytinen, K., Boland, R.J., Berente, N.: The Next Wave of Digital Innovation. Opportunities and Challenges: A Report on the Research Workshop 'Digital Challenges in Innovation Research'. SSRN Electronic Journal (2010)
14. Galanakis, K.: Innovation process. Make sense using systems thinking. Technovation 26, 1222–1232 (2006)

15. Pfeffer, J., Sutton, R.I.: Evidence-Based Management. Decision Making. Harvard Business Review 84, 62–72 (2006)
16. Osterwalder, A., Pigneur, Y.: Business model generation. A handbook for visionaries, game changers, and challengers. Wiley&Sons, New York (2010)
17. Gassmann, O., Frankenberger, K., Csik, M.: The business model navigator. 55 models that will revolutionise your business. Pearson Education Limited, Harlow, England, New York (2014)
18. Mathis, K., Felix Köbler: Data-Need Fit–Towards data-driven business model innovation. Proceedings of the ServDes. 5th Service Design and Innovation conference 125 (2016)
19. Forrester, J.W.: System dynamics - the next fifty years. System Dynamics Review 23, 359–370 (2007)
20. Sterman, J.D.: Business dynamics. Systems thinking and modeling for a complex world. McGraw-Hill Higher Education, Boston (2000)
21. Shafer, S.M., Smith, H.J., Linder, J.C.: The power of business models. Business Horizons 48, 199–207 (2005)
22. Massa, L., Afuah, A., Tucci, C.L.: A Critical Assessment of the Business Model and the Business Model Innovation Literature. Working Paper. Academy of Management Annals 11, 73–104 (2017)
23. Amit, R., Zott, C., Massa, L.: The Business Model: Recent Developments and Future Research. Journal of Management 37, 1019–1042 (2011)
24. Chesbrough, H., Rosenbloom, R.S.: The role of the business model in capturing value from innovation. Evidence from Xerox Corporation's technology spin-off companies. Industrial and Corporate Change 11, 529–555 (2002)
25. Hedman, J., Kalling, T.: The business model concept. Theoretical underpinnings and empirical illustrations. European Journal of Information Systems 12, 49–59 (2003)
26. Osterwalder, A., Pigneur, Y.: An eBusiness Model Ontology for Modeling eBusiness. Proceeding of the 15th Bled Electronic Commerce Conference (2002)
27. Morris, M., Schindehutte, M., Allen, J.: The entrepreneur's business model. Toward a unified perspective. Journal of Business Research 58, 726–735 (2005)
28. Teece, D.J.: Business Models, Business Strategy and Innovation. Long Range Planning 43, 172–194 (2010)
29. Massa, L., Tucci, C.L.: Business Model Innovation. In: Dodgson, M., Gann, D.M., Phillips, N. (eds.) The Oxford Handbook of Innovation Management, 20, p. 18. Oxford University Press, Oxford (2013)
30. Mitchell, D., Coles, C.: The ultimate competitive advantage of continuing business model innovation. Journal of Business Strategy 24, 15–21 (2003)
31. Woerner, S.L., Wixom, B.H.: Big data. Extending the business strategy toolbox. Journal of Information Technology 30, 60–62 (2015)
32. Foss, N.J., Saebi, T.: Fifteen Years of Research on Business Model Innovation. Journal of Management 43, 200–227 (2016)
33. Doz, Y.L., Kosonen, M.: Embedding Strategic Agility. Long Range Planning 43, 370–382 (2010)
34. Rumble, R., Mangematin, V.: Business Model Implementation: The Antecedents of Multi-Sidedness. Business Models and Modelling 33, 97–131 (2015)
35. Chesbrough, H.: Business Model Innovation. Opportunities and Barriers. Long Range Planning 43, 354–363 (2010)
36. Osterwalder, A., Pigneur, Y., Tucci, C.L.: Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of the Association for Information Systems 16, 1–25 (2005)

37. Forrester, J.W.: Industrial Dynamics - After the First Decade. Management Science 14, 398–415 (1968)
38. Shepherd, S.P.: A review of system dynamics models applied in transportation. Transportmetrica B: Transport Dynamics 2, 83–105 (2014)
39. Moellers, T., Bansemir, B., Pretzl, M., Gassmann, O.: Design and Evaluation of a System Dynamics Based Business Model Evaluation Method. In: Maedche, A., Jan vom Brocke, Hevner, A.R. (eds.) Designing the Digital Transformation, pp. 125–144. International Conference on Design Science Research in Information Systems DESRIST 2017: (2017)
40. Abdelkafi, N., Täuscher, K.: Business Models for Sustainability From a System Dynamics Perspective. Organization & Environment 29, 74–96 (2015)
41. Casadesus-Masanell, R., Ricart, J.E.: Competing through Business Models. Working Paper No. 713. Barcelona, Spain (2007)
42. Yin, R.K.: Case study research. Design and methods. Sage, Thousand Oaks, CA (2012)
43. Briscoe, G. and Mulligan, C.: Digital Innovation: The Hackathon Phenomenon. Creativeworks London Working Paper No. 6, http://www.creativeworkslondon.org.uk/wp-content/uploads/2013/11/Digital-Innovation-The-Hackathon-Phenomenon1.pdf (Accessed: 27.10.2017)
44. Mohajer Soltani, P., Pessi, K., Ahlin, K., Wernered, I.: Hackathon – a method for Digital Innovative Success: A Comparative Descriptive Study. Proceedings of the 8th European Conference on Information Management and Evaluation, ECIME 2014 8, 367–373 (2014)
45. Trainer, E.H., Kalyanasundaram, A., Chaihirunkarn, C., Herbsleb, J.D.: How to Hackathon: Socio-technical Tradeoffs in Brief, Intensive Collocation. Proceedings of the 2016 ACM Conference on Computer-supported Cooperative Work & Social Computing 19, 1116–1128 (2016)
46. Warner, J., Guo, P.J.: Hack.edu. Examining How College Hackathons Are Perceived By Student Attendees and Non-Attendees. In: Tenenberg, J., Chinn, D., Sheard, J., Malmi, L. (eds.) ACM International Computing Education Research conference (ICER), pp. 254–262 (2017)
47. Zorn, T.: Designing and conducting semi-structured interviews for research. Waikato Management School. Waikato Management School, Hamilton, New Zealand (2008)
48. Osella, M., Ferro, E.: Eight Business Model Archetypes for PSI Re-Use. “Open Data on the Web” Workshop. London (2013)
49. Kim, C., Mauborgne, R.: Blue Ocean Strategy. If you read nothing else on strategy, read these best-selling articles. Harvard Business Review 82, 71–80 (2004)
50. Casadesus-Masanell, R., Ricart, J.E.: How to Design a Winning Business Model. Harvard Business Review 89, 100–107 (2011)
51. Wernerfelt, B.: A Resource-Based View of the Firm. Strategic Management Journal 5, 171–180 (1984)
52. Winter, S.G.: Understanding dynamic capabilities. Strategic Management Journal 24, 991–995 (2003)

Business Model Innovation and Stakeholder: Exploring Mechanisms and Outcomes of Value Creation and Destruction

Sebastian Hermes[1], Markus Böhm[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Department of Informatics, Munich, Germany
{sebastian.hermes, markus.boehm, krcmar}@in.tum.de

Abstract. Given the objective of the focal firm to generate value for stakeholders, this research aims at assessing mechanisms and outcomes for value creation and destruction between business model innovation (BMI) and stakeholders. To achieve this goal, we conduct a systematic literature review and apply grounded theory as coding scheme. Taking frequent mechanisms and outcomes into account, we construct a conceptual framework and pioneer theory building. As main result, we identify BMI creating economic return for third parties and product/service access for customers. Both outcomes are based on the mechanism of altering resources and processes. In contrast, analyzing stakeholder's main influence, we find management creating strategic orientation by providing know-how. Our research agenda emphasizes the design of BMI from an ecosystem perspective and the destructive consequences of BMI. While the ecosystem level of analysis provides new insights into the concept, investigating negative impacts contributes to a more holistic understanding of BMI.

Keywords: Business model innovation, stakeholder theory, literature review, grounded theory, theory building

1 Introduction

The concepts of business models (BMs) and, more recently, BMI have become of increasing interest for scholars in recent years [1-3]. While BMs usually relate to firm-level value creation and capture [4], BMI also scrutinizes the novelty in value proposition as well as the logical and structural reorganization of firms [5]. The present paper defines BMI as a "search for new business logics of the firm and new ways to create and capture value for its stakeholders" [6], because it emphasizes the importance of an ecosystem perspective. One of the approaches to BMI recommended by Chesbrough [7] is to orient the firm towards an open BM. The concept of openness in BMs is viewed as being both innovative and cost-effective [8], which stresses the virtue of value creation and value capture when cooperating with external stakeholders. In addition, Tankhiwale [9] identifies that pressures from external stakeholders and regulations are often the drivers of BMI. Further reasons to involve stakeholders in the innovation process stem from managing conflicting objectives between internal and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

external stakeholders [10, 11], sensing new business opportunities [1], aligning and internalizing inter-organizational cognitions [12], strengthening a focal value proposition [13], and sustaining competitive advantage and profitability. Thus, some authors suggest that the objective of the firm is to generate value in different ways for different stakeholder groups [14]. Focusing on stakeholder theory is therefore vital to understand the emergence and consequence of BMI. The stakeholder-oriented approach becomes also relevant in the age of digital transformation as organizational boundaries are dispersing and the processes of value creation and capture are evolving from bidirectional to multidirectional, from centralized to decentralized, and from closed to open. As a consequence, stakeholders can be involved by applying open innovation approaches like idea communities [15] or idea competitions [16], but also through merger and acquisitions, joint development agreements, or inter-organizational negotiations [12]. To date, limited attention has been given to the reciprocal relationship between BMI and stakeholders despite the acknowledged influence stakeholders can exert on an organization's BM [9] and despite the fact that firms are reacting to innovations instead of driving them [17]. More specifically, Foss and Saebi [3] as well as Aspara, Lamberg, Laukia and Tikkanen [12] identify the need to examine the initiatives exerted on BMI by stakeholders while Spieth, Schneckenberg and Ricart [1] perceive the integration of stakeholders into the BMI process requiring further investigation. However, such fundamental questions are currently not systematically outlined, addressed, and answered. We are therefore providing a starting point with the present paper, which aims to contribute to the development and refinement of BMI by using a stakeholder lens [2, 18]. We determine the need for a more comprehensive view and assessment of value creation and destruction in a focal firm's ecosystem during the BMI process. Hence, the paper investigates what outcomes of value creation and destruction occur during BMI and the intervention of specific stakeholder groups. The outcomes are analyzed from a BMI perspective on the one side and from a stakeholder perspective on the other side. In addition, we present latent mechanisms pursued by each entity to achieve either value creation or destruction. Revealing these mechanisms is particular important to better describe and explain how value was created or destroyed [19]. The purpose of this paper is therefore to review current literature on the reciprocal relationship between BMI and stakeholders, evaluate them, and outline avenues for future research. While reviewing, synthesizing, and structuring current literature, we are guided by the following three research questions:

1. *Which outcomes does BMI have for stakeholders?*
2. *Which outcomes do stakeholder interventions have for BMI?*
3. *Which mechanisms account for the outcomes?*

2 Related Work

2.1 Business Models and Business Model Innovation

Although a focus of attention, the concept of BMs is "a slippery construct to study" [6]. Several frameworks of BMs have been seen in the literature so far [20-23]. A consensus is evolving to conceptualizing BMs as a holistic description and architecture of how value is created, delivered, and captured [24-26]. Thus, emphasizing the importance of integrating the perspective of stakeholders [27]. While interest in BMs is several decades old, the notion of BMs as distinct object of innovation was initially discussed in 2003 by Mitchell and Coles [28]. According to Zott, Amit and Massa [2], BMI can be characterized as a new dimension of innovation setting itself apart from process, product, and organizational innovation. Hence, giving rise to novel approaches for incremental or radical innovation of entire value chains, enabling competitive advantage and superior performance [29]. Due to the lack of construct clarity in BMs [30], it is not surprising that similar conclusions have been made with regard to the definition of BMI. However, various literature reviews attempt to categorize BMI research in unique streams paving the way for granular construct agreement [1, 3, 18]. For instance, Schneider and Spieth [18] present three major research streams: Prerequisites of conducting BMI, elements and processes of BMI, and results achieved through BMI. Building on these findings, Foss and Saebi [3] systematically investigate concepts, processes, outcomes, and consequences of BMI. This paper contributes not only to the research gaps of examining antecedents, outcomes, and boundary conditions of BMI as discussed by Foss and Saebi [3], but also to the effects and enablers of BMI since organizations often innovate their BMs as a reaction to changes in their environment [18].

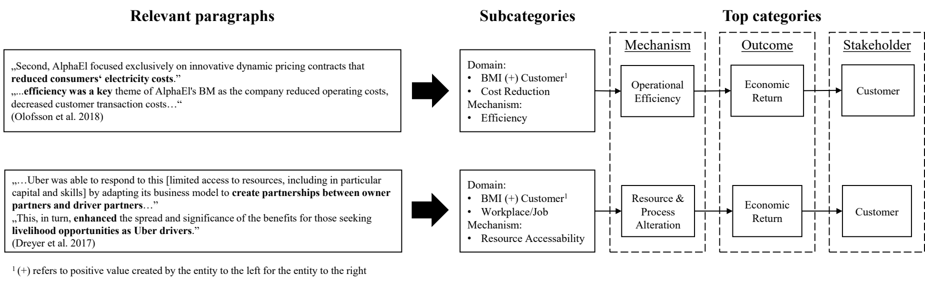
2.2 Stakeholder Theory in Business Model Research

According to Donaldson and Preston [31], stakeholder theory has turned into a major research stream in management literature. The concept is also widely recognized across different domains and becomes an increasingly important perspective for investigating BMs [32]. Freeman and Reed [33] define stakeholder as "any identifiable group or individual who can affect the achievement of an organization's objectives or who is affected by the achievement of the organization's objectives" and may be either primary (impacting the firm directly) or secondary (influencing the firm indirectly via primary stakeholders). Besides, stakeholders can be differentiated into internal and external stakeholders. While internal stakeholders include for example employees and top management teams, external stakeholders refer mainly to customers, users, suppliers, or universities [34]. Another well-established method to categorize stakeholders refers to arraying stakeholders on a power versus interest grid [35]. Freeman and Reed [33] argue that the responsibility for evaluating and mapping stakeholders lies at the top management level. Various researchers combining stakeholder theory and BMI agree on this perspective and regard the integration of stakeholders as a managerial task as well [36, 37]. Integrating customers is especially seen as a key activity for BMI. We infer from current literature that active stakeholder management is highly relevant to BMI research and that this implies developing strategies about when to integrate whom in which phase of BMI.

3 Design and Classification Paradigm of the Literature Review

Literature reviews are a well-known and rigorous approach to collect existing knowledge within an area of interest and to outline former research [38]. We found a descriptive review approach most appropriate for the present stage of this research [39]. We have therefore targeted three prominent online databases: Scopus, Web of Science, and EBSCOhost. Following the search terms of Foss and Saebi [3], we conducted title, keyword, and abstract searches across all three databases with the following query: (stakeholder OR partner* OR "Special interest groups" OR "Open Innovation") AND ("Business Model Innovation" OR "Business Model reinvention" OR "Business Model renewal" OR "Business Model dynamics" OR "Business Model transformation" OR "Business Model evolution") AND (effect* OR influenc* OR affect* OR impact*). The search identified a total of 101 articles. Following a staged selection process [40], the articles in the database were then scanned and filtered in two stages. The first stage involved removing duplicates as well as scanning titles and abstracts for apparently irrelevant articles. This stage of filtering excluded for example those articles that addressed the phenomena of new BMs instead of innovating an existing one or those articles that relied on the wording “partner” instead of describing the stakeholder they refer to in more detail. A total of 25 articles remained in the database. The second stage involved manually analyzing each article’s full text and including those articles that touched on the phases and components of BMI as well as distinct stakeholder specifications and precise value creation and destruction descriptions. By the end of this stage two articles were discarded, resulting in 23 remaining articles. In addition, we conducted a backward and forward search as recommended by Levy and Ellis [41]. We therefore reviewed all cited and citing papers of the 23 articles. We identified 10 additional articles, and therefore 33 peer-reviewed articles form the basis of the review in this paper. To systematically reveal and investigate academic insights on the reciprocal relationship of BMI and stakeholders, we developed a literature coding scheme. Figure 1 provides a small extract of the coding scheme.

Figure 1. Exemplary extract of the coding scheme



The extraction of insights was guided by the research questions raised earlier in this paper. In order to comply with our research aim, coding occurred on a textual level instead of categorizing the papers in general. Hence, an “open - axial - selective” approach informed by grounded theory [42] was adopted to identify the categories used for literature analysis. Such conventional and explorative content analysis has been recommended as a rigorous method for reviewing literature [43] and described as less confirmative than direct or summative approaches [44]. We assigned therefore specific subcategories to relevant paragraphs of each paper and then synthesized them into more generic top categories.

4 Descriptive Analysis

The 33 articles investigated account for a total of 319 subcategories. These split into 164 subcategories for mechanisms and 155 subcategories for outcomes. While the subcategories for the mechanisms converge into 13 top categories, 11 top categories emerge for the outcomes. The general focus has been on value creation and less on value destruction as destructive mechanisms and outcomes account for merely 79 subcategories altogether. It is noteworthy that the studies of Hienerth, Keinz and Lettl [45] and Olofsson, Hoveskog and Halila [46] make up the highest numbers of subcategories. While Olofsson, Hoveskog and Halila [46] explore BMI driven by sustainability issues at a social enterprise, Hienerth, Keinz and Lettl [45] focus on the implementation process of user-centric BMs. Both articles emphasize information and communication technology (ICT) as enabler and driver for digital transformation, which can act as antecedent for BMI [3]. However, BMI does not necessitates using ICT, in contrast, changing the logic of a firm can be achieved by different means [3]. The finding of ICT as trigger for digital BMs is also highlighted by most of the remaining articles [e.g. 47]. Moreover, the topic of sustainability appears to be another important unit of analysis as it is often mentioned as goal or purpose of BMI [e.g. 48]. The vast majority of articles have been published either in the areas of technology, innovation and entrepreneurship or in business administration literature. Around one fourth of articles stem from engineering and organization studies. The remaining articles are allocated to areas of sustainability, strategy, production, finance, and marketing. Interestingly, no article originates in information systems research despite the significance given to ICT and digital transformation in context of BMI. Further characteristics about the articles considered are illustrated in table 1.

Table 1. Descriptive analysis of articles considered

Paper Type		*VHB Ranking*		*Publication Year*		*Methodology*	
Journal	30	A	1	2018	4	Theory Paper	2
Conference	3	B	14	2017	6	Single Case Study	12
		C	8	2016	4	Multiple Case Study	10
		n.a.	10	2015	5	Regression Analysis	6
				2014	5	Structural Equation Model	1
				2013	3	Mixed Methods	2
				2012	2		
				2011	1		
				2010	2		

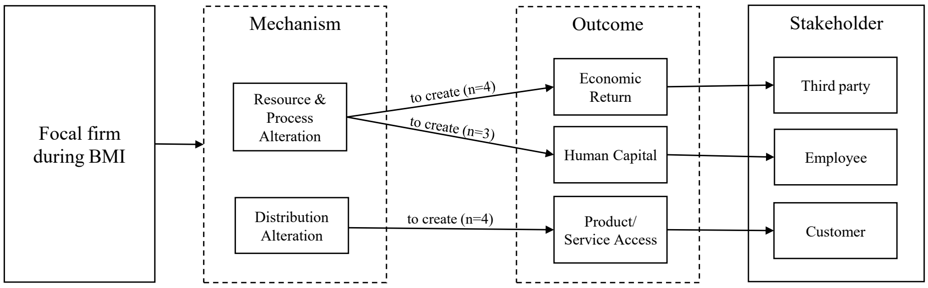
5 Towards a Conceptual Framework

We are now aiming to conceptualize the field, which might be a first attempt towards theory building [49]. Meredith [50] calls this a philosophical conceptualization, which in this case is based on reading the papers repeatedly. Since our goal is not only to describe the phenomenon accurately (outcome) but also to explain how it occurs (mechanism) and to whom (focal firm or stakeholder), we draw on the concept of context-mechanism-outcomes (CMO) pattern configuration. According to Linsley, Howard and Owen [19], "a CMO configuration is a proposition stating what it is about an initiative that works, for whom and in what circumstances." In this paper, context refers to BMI and stakeholder intervention while mechanisms and outcomes are investigated in order to develop an in-depth understanding about the reciprocal relationship between BMI and stakeholder intervention. Thus, we extracted according configurations only if the mechanism-outcome-stakeholder configuration had been identically mentioned by at least three articles. Doing so allows rigorous conceptual deduction of the cautiously proposed framework. The mechanisms and outcomes used to develop the framework stem from the synthesized top categories. The results are depicted in figure 2 and 3 and will be further explored in the next sections.

5.1 Business Model Innovation and Value Creation

As initial step, we identify the mechanisms used and the outcomes triggered by BMI to create value for particular stakeholder groups. On the one hand, we recognize how *altering resources and processes creates economic return for third parties.* Berti and Casprini [51] for example describe how an airport's processes changed towards offering extra-aviation activities. Thus, enabling shopping malls, parking providers, and restaurants to build flourishing businesses at the airport. On the other hand, we notice that *resource and process alteration also benefits employees in form of fostering their human capital*. While Aspara, Lamberg, Laukia and Tikkanen [12] stress how Nokia's business model transformation led to the selection of business that enhanced the development of corporate human resources, Carayannis, Sindakis and Walter [48] mention that the organizational transition towards servitization encouraged employees to adopt new skills and knowledge. Next, we present our findings about the *alteration of distribution channels and its positive influence on the customer's access to products and services*. By investigating how an original equipment manufacturer innovated its BM towards becoming an own brand and product developer, Carayannis, Sindakis and Walter [48] observed an expansion of direct sales from wholesalers to single retailers. Hence, allowing additional customers in the value chain to access its products. Ghezzi, Cavallaro, Rangone and Balocco [17] find a similar effect studying BMI in the context of mobile portals and their shift to application stores. In mobile portals, the customer's access is limited to the operator's portal. The portal represents the sole interface through which end customers obtain content and service offers. By engaging in the application creation and distribution paradigm, the focal firm permits higher openness and independence to third parties, providing users broader product and service choices. Moreover, the firm integrates application developers as a new customer group and transforms its business model into a two-sided market. Figure 2 illustrates the mechanisms used and outcomes triggered by BMI to create value for particular stakeholder groups.

Figure 2. Business model innovation and value creation for stakeholder



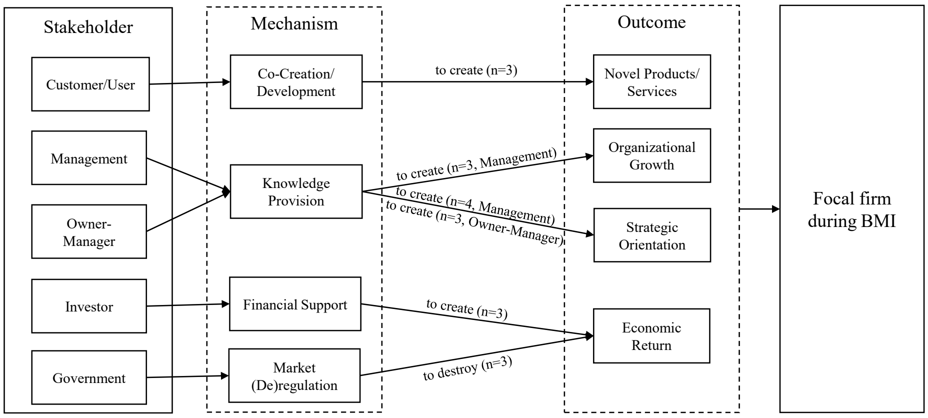
5.2 Stakeholder Intervention and Value Creation

This section describes how the mechanisms used and the outcomes triggered by different stakeholder groups enhance the BMI of the focal firm. First, we present our findings about the *beneficial effect of customers and users engaging in co-creation in new product or service development*. In their multiple case study, Hienerth, Keinz and Lettl [45] investigate the success factors of involving users in core business processes. Doing so, they report that the company LEGO engaged continuously with its users in co-creation resulting in the launch of the LEGO Factory platform - now called LEGO Ideas. The authors observed the same pattern at the company Coloplast, which integrates users in order to co-create new products with the development staff. Interestingly, the companies in both cases relied on IT tools to improve their co-creation processes since these IT tools facilitated large-scale user interaction and effective information collection. Accordingly, Kohler [52] delineates how various integrator platforms offer products that are co-created by the crowd ranging from t-shirts sold on Threadless to cards sold on Minted. In case of product platforms, the author identifies a similar co-creation procedure and refers to Apple's IOS and Google's Android ecosystem. Both companies allow users to develop and distribute their apps on top of their platforms. Hence, crowd members co-create new products or services with platform providers. Secondly, we discuss how *management's provision of knowledge creates organizational growth* for the focal firm during BMI. Abebe and Myint [53] identify that board members facilitate BMI adoption because they provide valuable information on changes in the external environment. Accordingly, management can positively contribute to firm performance by providing valuable and relevant external information. Similarly, Guo, Zhao and Tang [54] provide statistical support for the positive effect of top managers' human capital on BMI. More specifically, the authors show how combining top managers' managerial skills and managerial ties might enable the focal firm to capitalize on existing opportunities, whereas top managers' entrepreneurial skills can guide the focal firm to convert information and knowledge acquired through managerial ties into new business or product opportunities. Thirdly, we outline our findings on the *positive influence between managers' or owner-managers' provision of knowledge and the focal firm's strategic orientation*. In addition to enhancing organizational growth, Abebe and Myint [53] also show that larger boards can positively contribute to firm strategy since their extensive knowledge improves the quality of strategic decisions. Hence, management teams provide the human capital necessary to adopt new innovations in the marketplace. By analyzing Nokia's corporate BM transformation, Aspara, Lamberg, Laukia and Tikkanen [12] describe how top managers seek to retain or renew existing BM elements. While a corporate crisis led top managers to decide on changing Nokia's BM to a new, more legitimate corporate recipe, it was top management's knowledge about strategic fit and complementarity product that enabled the firm to reformulate its strategic positioning. Regarding owner-managers, Olofsson, Hoveskog and Halila [46] state that the social vision and the business experience of the founder were especially crucial factors contributing to the success of the focal firm. For example, the founder introduced new marketing ideas like environmentally certified electricity, which attracted new customers. Interestingly, once the founder resigned, the firm experienced a strategic identity crisis to some degree. Additionally, Velu and Jacob [55] argue that entrepreneurs that are also managers comprise a more holistic understanding of the business and more comprehensive insights about internal and external environments. Therefore, owner-managers enable the systemic and strategic change that BMI demands. Finally, we elaborate how *investors create economic return by providing financial support* to the focal firm engaging in BMI. While Berti and Casprini [51] describe that investors became an important source of revenues by acquiring company equity, Olofsson, Hoveskog and Halila [46] scantily state that the investor's financial support was critical to the success of the sustainability-driven firm investigated. Moreover, Demil and Lecocq [21] elucidate how one major investment enabled an English football club to build new infrastructure and improve personnel training. These developments permitted the football club to counter negative impacts resulting from legal rulings.

5.3 Stakeholder Intervention and Value Destruction

The following section depicts how *market regulations and deregulations implemented by the government destroy economic return*. In their search of dynamic consistency during BMI, Demil and Lecocq [21] illustrate how regulation reduced revenues and deregulation increased costs. The governmental regulation was grounded in the Taylor report and forced an English football club to reduce the capacity of its stadium by almost 50 percent. As a consequence, the club was facing the prospect of regular losses by the end of the 90s. In contrast, the Bosman ruling relaxed the existing transfer system and relieved football players from their preposterous contractual status. This deregulation facilitated competition for the best players between European clubs which raised both salaries and transfer fees. Similarly, Sosna, Trevinyo-Rodríguez and Velamuri [56] report that the deregulation of the Spanish dietary products market eased the government registration of products. Hence, the focal firm had to contend for shelf space against incumbents, who competed on brand strength and product range, and against new firms competing on price. Figure 3 illustrates how the mechanisms used and the outcomes triggered by different stakeholder groups create and destroy value for the focal firm engaging in BMI.

Figure 3. Stakeholder and value creation and destruction for the focal firm



6 Future Research

6.1 Designing Business Model Innovation from an Ecosystem Perspective

Our review revealed that all studies focused on BMI from a firm-centered, inside-out perspective, neglecting network relationships [10, 51, 57, 58]. Hence, future research can gain additional insights from applying an ecosystem perspective that goes beyond the dyadic stakeholder-firm relationship. Spanning organizational and bilateral borders does not only enhance our understanding of the consequences of BMI, but it also reveals a new context to which the purpose of BMI can be aligned to. Instead of striving to create value solely for the firm or different stakeholder groups, BMI can be designed to propose and create value for the entire ecosystem it operates in. We argue that adopting such a holistic approach alters the purpose of BMI towards more sustainable business practices. The underlying reasoning is two-folded. First, we draw on general equilibrium theory [59] and derive that value creation on the one side leads to value destruction on the other side of the ecosystem. However, as is typical for biological ecosystems, once one side of the ecosystem suffers it also affects the other side of the ecosystem. Destroying value in one part of the ecosystem will therefore sooner or later affect the firm initiating the value destruction in the first place. Secondly, we feel that the understanding of this circular independency leverages preventive activities. Thus, firms applying the ecosystem level of analysis to BMI will adopt more sustainable business practices. Theoretical contributions can be made in two ways. First, to the position-based view of the firm as the company adjusts its position in response to environmental and market forces following an outside-in perspective. Second, to the ecosystem concept as the company aligns its structure, processes, and activities towards proposal and creation of value for a multilateral set of stakeholders and ecosystem actors [13]. Building on the above reasoning, we propose to investigate the following research questions: Who to design BMI for and for which purpose(s)? When to integrate which kind of ecosystem actor to achieve the selected purpose(s)? How to design BMI to create and maintain sustainable business practices? How to govern sustainable business practices? How to incentivize direct and indirect stakeholder to participate in sustainable business practices? Studying these questions can provide practitioners with novel concepts on how to build sustainable business growth and enhance firm survival.

6.2 Exploring Value Destruction of Business Model Innovation

In our analysis of existing literature, we identified that the concept of value destruction as a consequence of BMI is being under-researched. Current research efforts do scratch the surface of value destruction, but hardly manage to investigate it in more detail. In cases where they do explore value destruction, it is solely in terms of how stakeholders affect BMI, but not the other way around. For example, Holm, Günzel and Ulhøi [47] mention how several cases of value destruction impede BMI. The cases range from complying with third-party standards due to cooperation with sales intermediaries to competing with users due to new ICT involving users in value creation and diffusion. However, they miss to explore the underlying mechanisms and impacts more thoroughly. In contrast, research on the government as destructive trigger for BMI has been widely investigated so far. For example, Demil and Lecocq [21] illustrate how governmental regulation reduced revenues and how deregulation increased costs during the phase of BMI. We argue that the concept of value destruction provides avenues for fruitful research, especially when investigating how BMI destroys value for the actors in the ecosystem. At present, research is concentrated on only one side of the coin, value creation, but neglects to explore value destruction as the other, as important side of the coin. Engaging with the proposed concept provides additional insights on the emergence, mechanisms and consequences of value destruction. Therefore, contributing to the other, the negative side of BMI. Following this concept helps not only to understand how BMI affects primary stakeholders, but also how it impairs secondary actors in the ecosystem. We feel that negative externalities in particular provide interesting phenomena to explore in future endeavors. Therefore, we are calling for exploration of the following research questions: How do customers, suppliers, complementors, competitors etc. inhibit firms from aligning their BMI with ecological, societal, and financial goals? How and to which degree do the negative externalities of BMI affect stakeholders that are not part of the firm's direct network? Evidence and motivation for negative externalities can be observed at Uber and Airbnb [60]. At Airbnb for example, hosts do not pay lodging taxes, therefore municipalities lose tax revenues and hotels suffer from unfair competition. Moreover, landlords find their long-term tenants turning into short-term landlords, unjustly enriching themselves and skirting rent stabilization laws. Another group of indirect stakeholders, neighborhoods, claim to be overrun by visitors bringing noise, trash and traffic. In sum, the negative externalities of Airbnb can decrease the amount of housing and increase renting prices [61]. Consequently, homes for residents who work within the city, participate at votes, build families, or simply have no other place to go, are being diminished. During the investigation of externalities, research should not only focus on case studies of constructive BMI; insights from destructive BMI can enhance the field and provide new perspectives. Patterns for the design and strategies for the governance of sustainable business development could emerge in multiple-case studies of constructive and destructive BMI and their impact on the economic ecosystem.

7 Limitation and Conclusion

Several limitations affect the results of our study. First, the literature search might not cover all relevant studies due to the choice of keywords. For example, alternative terms for the concept of stakeholder such as partner, competitor, employee, government etc. might yield further relevant articles. Second, the applied coding process simplifies the results of the studies to make them comparable. Similar subcategories were assigned to more generic top categories. In the course of this process, some insights might have been lost and may not be represented in our results. To conclude, we uncovered latent mechanisms and outcomes of value creation and destruction by applying an open, axial, and selective coding approach to synthesize implicit insights of the 33 articles identified by our keyword search. Abstracting from individual findings, we attempted to construct a conceptual framework relating prevalent mechanisms to specific outcomes and stakeholders, hence, clarifying the reciprocal relationship of BMI and stakeholders. We identified two relationships as main results on how BMI creates value for stakeholders. First, BMI creates economic returns for third parties by altering resources and processes. Second, BMI creates product/service access to customers by altering resources and processes as well. Reversing the direction of impact to stakeholders influencing BMI, the main result emerges from management creating strategic orientation for BMI by providing their knowledge. Last, we outlined potential avenues for future research. We recommend to study the design of BMI from an ecosystem perspective. The new level of analysis will provide further insights into the concept of BMs and is highly relevant in practice. Moreover, we think that future research needs to explore the destructive side of BMI. Investigating the negative consequences of BMI will contribute to a more holistic understanding of BMI. By reviewing existing literature and deriving issues for future research, our study contributes to information systems literature in several ways. First, we provide an overview on research related to the beneficial and destructive impacts between BMI and stakeholders. The overview highlights new insights that were previously incorporated implicitly in the literature. Second, we summarize mechanisms and outcomes related to value creation and destruction across all studies. In doing so, we identify and illustrate the key concepts currently being touched on by scholars in the field of BMI. Third, we expand existing theory on BMI by identifying and explaining those antecedents of BMI which Foss and Saebi [3] call stakeholder demands. Addressing their proposed gap number two, we provide insights about internal and external stakeholder demands and illustrate what Aspara, Lamberg, Laukia and Tikkanen [12] call "initiatives of other stakeholders than managers (or investors)." Moreover, we contribute to theory on outcomes of BMI by taking an ecosystem perspective. Instead of investigating what outcomes BMI has for the focal firm, we explain what outcomes BMI has for its stakeholders. Fourth, we derive specific issues for future research that are rooted in existing research but show how our understanding of BMI and its design can be enhanced. Finally, our study is relevant for practice by laying out which impacts practitioners need to consider when engaging in BMI. The issues we identified will prove to be useful in practice and will further advance the applicability of the scientific findings during BMI.

8 Acknowledgements

This research is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG) as part of the 'Collaborative Research Center 768: Managing cycles in innovation processes – Integrated development of product service systems based on technical products' (TP C1). For helpful comments and suggestions, the authors would like to thank the anonymous reviewers and the editors.

References

(* included in the literature analysis)

1. Spieth, P., Schneckenberg, D., Ricart, J.E.: Business model innovation–state of the art and future challenges for the field. R&d Management 44, 237-247 (2014)
2. Zott, C., Amit, R., Massa, L.: The Business Model: Recent Developments and Future Research. J Manage 37, 1019-1042 (2011)
3. Foss, N.J., Saebi, T.: Fifteen Years of Research on Business Model Innovation: How Far Have We Come, and Where Should We Go? J Manage 43, 200-227 (2017)
4. Weking, J., Hein, A., Böhm, M., Krcmar, H.: A Hierarchical Taxonomy of Business Model Patterns. Electronic Markets In Press, (2018)
5. Weking, J., Stöcker, M., Kowalkiewicz, M., Böhm, M., Krcmar, H.: Archetypes for Industry 4.0 Business Model Innovations. 24th Americas Conference on Information Systems (AMCIS 2018), New Orleans, LA, USA (2018)
6. Casadesus-Masanell, R., Zhu, F.: Business model innovation and competitive imitation: The case of sponsor-based business models. Strategic Manage J 34, 464-482 (2013)
7. Chesbrough, H.W.: Open Business Models: How to Thrive in the New Innovation Landscape. Harvard Business School Press, Boston, MA, USA (2006)
8. Chesbrough, H.W., Appleyard, M.M.: Open innovation and strategy. California Management Review 50, 57-76 (2007)
9. Tankhiwale, S.: Exploring the interrelationship between telco business model innovation and the change in business process architecture. Journal of Telecommunications Management 2, 126-137 (2009)
10. Miller, K., McAdam, M., McAdam, R.: The changing university business model: a stakeholder perspective. R&D Management 44, 265-287 (2014)
11. Gassmann, O., Frankenberger, K., Sauer, R.: A Primer on Theoretically Exploring the Field of Business Model Innovation. TEBR The European Business Review (2017)

12. *Aspara, J., Lamberg, J.-A., Laukia, A., Tikkanen, H.: Corporate business model transformation and inter-organizational cognition: the case of Nokia. Long Range Planning 46, 459-474 (2013)
13. Adner, R.: Ecosystem as structure: an actionable construct for strategy. Journal of Management 43, 39-58 (2017)
14. Post, J.E., Preston, L.E., Sachs, S.: Managing the extended enterprise: The new stakeholder view. California management review 45, 6-28 (2002)
15. El Sawy, O.A., Kræmmergaard, P., Amsinck, H., Vinther, A.L.: How LEGO Built the Foundations and Enterprise Capabilities for Digital Leadership. MIS Quarterly Executive 15, 141-166 (2016)
16. Füller, K., Weking, J., Böhm, M., Krcmar, H.: Leveraging Customer Integration Experience: A Review of Influencing Factors and Implications. Communications of the Association for Information Systems (CAIS) In Press, (forthcoming)
17. *Ghezzi, A., Cavallaro, A., Rangone, A., Balocco, R.: On business models, resources and exogenous (dis) continuous innovation: evidences from the mobile applications industry. International Journal of Technology Management 68, 21-48 (2015)
18. Schneider, S., Spieth, P.: Business model innovation: Towards an integrated future research agenda. International Journal of Innovation Management 17, 1340001 (2013)
19. Linsley, P., Howard, D., Owen, S.: The construction of context-mechanisms-outcomes in realistic evaluation. Nurse Researcher 22, 28-34 (2015)
20. Johnson, M.W., Christensen, C.M., Kagermann, H.: Reinventing Your Business Model. Harvard Bus Rev 86, 57-68 (2008)
21. *Demil, B., Lecocq, X.: Business Model Evolution: In Search of Dynamic Consistency. Long Range Planning 43, 227-246 (2010)
22. Baden-Fuller, C., Mangematin, V.: Business models: A challenging agenda. Strategic Organization 11, 418-427 (2013)
23. Chesbrough, H.: Business Model Innovation: It's not just about technology anymore. Strategy & Leadership 35, 12-17 (2007)
24. Osterwalder, A., Pigneur, Y., Tucci, C.L.: Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of the Association for Information Systems 16, 1-25 (2005)
25. Shafer, S.M., Smith, H.J., Linder, J.C.: The power of business models. Business horizons 48, 199-207 (2005)
26. Teece, D.J.: Business Models, Business Strategy and Innovation. Long Range Planning 43, 172-194 (2010)
27. Zott, C., Amit, R.: Business model design: an activity system perspective. Long range planning 43, 216-226 (2010)
28. Mitchell, D., Coles, C.: The ultimate competitive advantage of continuing business model innovation. Journal of Business Strategy 24, 15-21 (2003)
29. Baden-Fuller, C., Morgan, M.S.: Business Models as Models. Long Range Planning 43, 156-171 (2010)
30. Fjeldstad, Ø.D., Snow, C.C.: Business models and organization design. Long Range Planning 51, 32-39 (2018)
31. Donaldson, T., Preston, L.E.: The stakeholder theory of the corporation: Concepts, evidence, and implications. Academy of management Review 20, 65-91 (1995)
32. Gassmann, O., Frankenberger, K., Sauer, R.: Exploring the Field of Business Model Innovation: New Theoretical Perspectives. Springer (2016)
33. Freeman, R.E., Reed, D.L.: Stockholders and stakeholders: A new perspective on corporate governance. California management review 25, 88-106 (1983)

34. Jonas, J.M., Roth, A.: Stakeholder integration in service innovation-an exploratory case study in the healthcare industry. International Journal of Technology Management 73, 91-113 (2017)
35. Bryson, J.M.: What to do when stakeholders matter: stakeholder identification and analysis techniques. Public management review 6, 21-53 (2004)
36. Harrison, J.S., Bosse, D.A., Phillips, R.A.: Managing for stakeholders, stakeholder utility functions, and competitive advantage. Strategic management journal 31, 58-74 (2010)
37. Gnatzy, T., Moser, R.: Scenario development for an evolving health insurance industry in rural India: INPUT for business model innovation. Technological Forecasting and Social Change 79, 688-699 (2012)
38. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly xiii-xxiii (2002)
39. King, W.R., He, J.: Understanding the role and methods of meta-analysis in IS research. Communications of the Association for Information Systems 16, 32 (2005)
40. Dybå, T., Dingsøyr, T.: Empirical studies of agile software development: A systematic review. Information and software technology 50, 833-859 (2008)
41. Levy, Y., Ellis, T.J.: A systems approach to conduct an effective literature review in support of information systems research. Informing Science 9, 181-212 (2006)
42. Glaser, B.G., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Transaction (1967)
43. Wolfswinkel, J.F., Furtmueller, E., Wilderom, C.P.: Using grounded theory as a method for rigorously reviewing literature. European journal of information systems 22, 45-55 (2013)
44. Hsieh, H.-F., Shannon, S.E.: Three approaches to qualitative content analysis. Qualitative health research 15, 1277-1288 (2005)
45. *Hienerth, C., Keinz, P., Lettl, C.: Exploring the nature and implementation process of user-centric business models. Long Range Planning 44, 344-374 (2011)
46. *Olofsson, S., Hoveskog, M., Halila, F.: Journey and impact of business model innovation: The case of a social enterprise in the Scandinavian electricity retail market. Journal of Cleaner Production 175, 70-81 (2018)
47. *Holm, A.B., Günzel, F., Ulhøi, J.P.: Openness in innovation and business models: lessons from the newspaper industry. International Journal of Technology Management 61, 324-348 (2013)
48. *Carayannis, E.G., Sindakis, S., Walter, C.: Business model innovation as lever of organizational sustainability. The Journal of Technology Transfer 40, 85-104 (2015)
49. Weick, K.E.: What theory is not, theorizing is. Administrative science quarterly 40, 385-390 (1995)
50. Meredith, J.: Theory building through conceptual methods. International Journal of Operations & Production Management 13, 3-11 (1993)
51. *Berti, C., Casprini, E.: When regulatory changes become a driver for business model innovation: An illustrative case in an Italian airport. International Journal of Organizational Analysis 26, 63-74 (2018)
52. *Kohler, T.: Crowdsourcing-based business models: how to create and capture value. California Management Review 57, 63-84 (2015)
53. *Abebe, M., Myint, P.P.A.: Board Characteristics and the Likelihood of Business Model Innovation Adoption: Evidence from the Smart Home Industry. International Journal of Innovation Management 22, 1850006 (1850028 pages) (2018)
54. *Guo, H., Zhao, J., Tang, J.: The role of top managers' human and social capital in business model innovation. Chinese Management Studies 7, 447-469 (2013)

55. *Velu, C., Jacob, A.: Business model innovation and owner–managers: the moderating role of competition. R&D Management 46, 451-463 (2016)
56. *Sosna, M., Trevinyo-Rodríguez, R.N., Velamuri, S.R.: Business model innovation through trial-and-error learning: The Naturhouse case. Long range planning 43, 383-407 (2010)
57. *Bowyer, D., Chapman, R.L.: Does privatisation drive innovation? Business model innovation through stakeholder viewpoints: the case of Sydney Airport 10 years post-privatisation. Journal of Management & Organization 20, 365-386 (2014)
58. *Wang, D., Guo, H., Liu, L.: One goal, two paths: How managerial ties impact business model innovation in a transition economy. Journal of Organizational Change Management 30, 779-796 (2017)
59. Hahn, F.: General equilibrium theory. The Public Interest 0, 123-139 (1980)
60. Edelman, B.G., Geradin, D.: Efficiencies and regulatory shortcuts: How should we regulate companies like Airbnb and Uber. Stan. Tech. L. Rev. 19, 293-326 (2015)
61. Interian, J.: Up in the air: Harmonizing the sharing economy through Airbnb regulations. BC Int'l & Comp. L. Rev. 39, 129-161 (2016)

Business Models for Internet of Things Platforms: Empirical Development of a Taxonomy and Archetypes

Daniel Hodapp[1], Gerrit Remane[2], Andre Hanelt[1], and Lutz M. Kolbe[1]

[1] University of Göttingen, Chair of Information Management, Göttingen, Germany
daniel.hodapp@stud.uni-goettingen.de
{andre.hanelt, lutz.kolbe}@wiwi.uni-goettingen.de
[2] FH Wedel, IT-Management, Wedel, Germany
gerrit.remane@fh-wedel.de

Abstract. A wide range of Internet of Things platform providers operate diverse business models to cater for the manifold requirements of the IoT. This paper contributes to a more precise understanding of IoT platforms as an essential building block of the IoT based on the characteristics of its business models. Even though research listing technological dimensions according to which IoT platforms differ, they have neither been systematically derived nor been linked to the business model concept. In turn, they lack descriptive power on the heterogeneous value creation mechanisms of the platform providers. Within our research, we first analyzed 195 IoT platforms and systemically developed a taxonomy allowing the characterization of IoT platform business models. Second, based on this taxonomy, we identified nine archetypes of IoT platform business models and illustrated typical combinations of business model characteristics. Equipped with such an understanding, practice and research can analyze existing IoT platforms more accurately.

Keywords: Internet of things, IoT platforms, Taxonomy, Business model, Archetypes

1 Introduction

There are 20 billion connected devices forecasted by 2020 [1]. For connecting all these heterogeneous physical products to the digital world, IoT platforms are essential. They act as a central backbone connecting the heterogeneous device landscape, different data sources, developers and end-users [2]. To give an illustrative case: The cloud company Amazon operates an IoT platform and provides the functionality to manage millions of devices. Other companies can build on these functions and develop IoT applications. For instance, Miovision, a company offering a variety of devices for measuring air quality and noise builds on Amazons platform to connect its physical devices and to develop applications that support cities in improving transport capacity, safety, efficiency and performance [3]. However, in order to take advantage of the economic opportunities more than 450 firms have entered the market for IoT platforms [4]. While all platforms are offered under an IoT related platform label, they vary considerably as they focus on different aspects of the IoT technology stack and correspondingly include diverse functionalities in their offerings, leading to different types of value creation [5].

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

So far, academia has highlighted different technological characteristics of IoT platforms. Some researchers stress the integrative role of IoT platforms and describe them as an abstraction service for smart objects, i.e. a digital representation of physical objects equipped with sensors, actuators and communication technology is available on the platform [6, 7]. Thereby, the IoT platform can be operated on the cloud level and provide the functionalities in a platform as a service manner or in a local environment running on a gateway or other devices[6]. In addition, some researchers include the role of end-users and characterize platforms as interaction enablers between end-users and smart objects by providing the necessary infrastructure [2]. Later on, the understanding of IoT platforms as instantiation of a digital platform emerged, i.e., a software-based system that provides core functionality shared by the interoperating modules [5, 8]. These core functionalities include integration, management and monitoring of smart objects, that can be utilized to build applications for the IoT [6]. In sum, IoT platforms can be understood as a specific type of digital platforms that are (i) operated in a cloud or local environment, (ii) enable the interaction between smart objects and end users (iii) by providing a core functionality to third party developer to support the development of modular applications (iv) on the basis of an abstraction service that is integrating underlying infrastructure and different data sources.

To advance the understanding of the IoT, existing technological considerations must be complemented by a business perspective [9]. Against this backdrop, we argue that a business model-focused view on IoT platforms is necessary, as it provides just the right perspective for relating rather technological aspects (e.g., platform openness) to essential business decisions (e.g., pricing model), thereby making these interdependencies explicit and thus manageable as part of a systematic business model design process. Considering this need, the paper strives to answer the following two research questions: *(1) How can instances of IoT platform business models be characterized?, (2) What are the IoT platform business model archetypes that can be differentiated?* To do so, we briefly provide the necessary background to digital business model frameworks before we proceed in four phases: (1) We develop a representative database of IoT platform providers. (2) We carry out a systemic literature review to identify publications that introduce specific dimensions along IoT platforms. (3) We deconstruct the business model of platform providers into a classification scheme that serves as a taxonomy and thus, a theory for analysis [10]. (4) Lastly, we empirically identify business model archetypes from this taxonomy.

2 Digital Business Model Frameworks

The term business model describes how an organization creates, delivers, and captures value [9]. In literature, the concept mainly serves as a classification basis for emerging value drivers, as performance indicator, or as foundation for business model innovation [11]. Several authors have introduced business model frameworks that provide design options for specific business model components e.g., [9]. This component-based view on the business model concept i.e., a system comprising a set of interacting components, is dominating the current scientific discussion [11]. However even though

an intense scholarly discourse, there is no consensus on what a business model actually is, nor on what specific set of relevant components are included[1] [12].
With respect to our research questions, we focus on frameworks that were developed explicitly for digital business. This is necessary as IoT platforms (as well as many other platforms) rely heavily on digital technologies to create value and can therefore be considered digital business models [13]. We identified several specific business model frameworks describing the generic dimensions of digital business models, most importantly [14], [15] and [16]. In addition, other authors have also introduced business model taxonomies for specific digital domains. These approaches are more precise as they not only describe the component and dimensions of digital business models in a certain domain, but also specify the potential characteristics for each of these dimensions. Therefore, they are not only useful for analyzing existing business models but are also a powerful tool for business model innovation [17]. Therefore the first objective of this research is to develop a business model taxonomy for IoT platforms.

3 Methodology

In prior publications, the taxonomy development was conducted for two reasons: first, to provide the fine-grained approach to re-structure a comparatively young research domain (e.g. 17), and second, to set the ground to extract archetypes (e.g., 18). We included central elements from these studies in our four-phase research design.

3.1 Phase 1

In the first phase, we created a database of IoT platforms that were operational between May and June 2018 following three steps: first, we combined a variety of sources, starting from IoT platforms mentioned in the literature review (54 platforms), continuing with a tag search for "IoT platform" in the world's largest tech database CrunchBase (221 platforms), and finally adding the platforms of the leading IoT database Postscapes (123 platforms). By pooling the sources, we obtain 398 platforms. Second, we excluded providers who occurred more than once, went bankrupt, got acquired, ceased operation or offered no market-ready product. As a result, we have a reduced data set of 237 IoT platforms. Third, to have a consistent understanding of an IoT platform, we excluded platforms that did not follow the introduced understanding of an IoT platform (see introduction). The database contains all the leading platforms according to multiple market analyses (e.g., [1]). Further commercial reports indicated the presence of more than 450 IoT platforms [4]. However, we consider our database with 195 operating platforms as a sufficiently representative sample of the overall IoT platform market to identify the most relevant business model dimensions as well as archetypical combinations of these dimensions.

[1] For further theoretical details on the business model concept, see [12]

3.2 Phase 2

In the second phase, a systematic literature search and analyzes was carried out [19]. To identify all potentially relevant publications, we conducted a keyword search using the databases ScienceDirect, IEEE, AISEL and Google Scholar in the title, abstract and keywords by using the following search strings[2]: "IoT", "Internet of things", "platform", "middleware", "cloud", "classification", "types", "typology", "taxonomy", "business model" resulting in an initial sample of 327 publications.

Table 1. Literature based dimensions

Author	Conceptual dimensions	Literature-based taxonomy dimensions
[20]	• Core capabilities • Platform openness	Context awareness, Device management, Interoperation, Platform portability, Security and privacy
[21]	• Core capabilities, • Platform integration • Partner system • Degree of support	Access, Architecture, Data store, Documentation, External integration, Fault tolerant, Finality, Languages, Last update, License, Load balancing, Memory, Official Companies, Programming, Replication
[22]	• Core capabilities • Platform openness	Analytic tool, API/library, Connection type, Database type, Data formats, Energy efficiency, Messaging protocol, Notification/alert, Open Source/commercial, Platform resource requirements, Programming Language, Security, Software Development Tool, Storage Space, Visualization
[23]	• Platform integration • Degree of support	Device Management, Conversion Security, Decentralization, Enterprise integration, Event-Management, Filtering & enhancement, Information, Maintenance, Technology abstraction, Visualization
[2]	• Device support • Application sales channel • Partner system • Degree of support	Data ownership, Data processing and sharing, Developer support, Ecosystem formation, IoT marketplace, Support of heterogeneous devices
[24]	• Core capabilities • Device support • Degree of support	Availability, Developer Friendly, Plug and Play, Provision of Support, Real time data, Scalability, Security & Privacy Provisioning, Solution Type, Storage of data
[25]	• Core capabilities • Device support • Transaction based revenues	API protocols, Application, Cloud model type, Data analytics, Data visualization, Device configuration, Real time data capture capability, Usage costs / developer costs
[26]	• Core capabilities • Platform openness	Integration to Cloud, Mostly Used applications, Security, Supporting Protocols, Type of Analytics

To further systematize the review process we proceeded in three steps. First, we screened the abstract with regard to our research question (sample size 147). Second, we solely included publications in peer reviewed journals or conferences to ensure scientific rigor (sample size 96). Third, we conducted an in depth analyzes were each publication had to contain technical or other business model related dimensions that support the description of IoT platform business models (sample size 7). We than conducted a forward-backward, resulting in sample of eight publications (table 1).

In the next step, we systematically analyzed the literature corpus and identified three limitations: (1) The literature based dimensions are not systematically derived, but were rather developed ad-hoc, and lack details regarding the origin of each dimension and the associated characteristics, resulting in lack of transparency. (2) The presented dimensions were mostly developed in a technical context and not linked to the business model concept, thereby neglecting important components of a complete business model. Possibilities to capture value remain largely unexplored. (3) The introduced taxonomy dimensions are only sporadically evaluated on a representative real world sample of IoT Platforms. In addition, we have derived eight conceptual dimensions

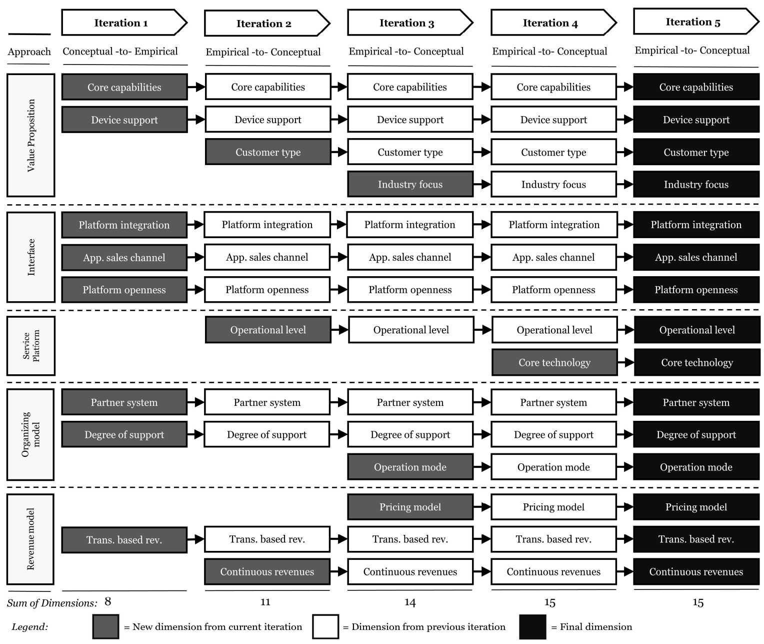
[2] Some contributions use platform [2], middleware [20] and cloud [21] interchangeable thus we included them in the key word search.

from the literature, which serve as a starting point for the taxonomy development and provide researchers with initial notions of how IoT platform business models differ.

3.3 Phase 3

In the third phase, a taxonomy of the IoT platform business models was systematically developed. The process was guided by the taxonomy development method of [27]. The method is rigorous, since the necessary steps and ending conditions are clearly defined and proven in numerous applications. In addition, the method enables the combination of already identified dimensions with new dimensions from the research of real objects.

Figure 1. IoT platform business model taxonomy development



In the first step of the process, we casted our meta-characteristic within the VISOR framework from [15], similar to [17], to assure that all relevant components of digital business models are covered. The VISOR framework is the most adequate one for decomposing the business model of IoT platform providers, as it emphasizes the importance of the customer interface, the central role of digital platforms, and the need to orchestrate complex digital ecosystems. The VISOR framework consists of five components: value proposition, interface, service platform, organizing model and revenue model. Second, our taxonomy development process is based on the eight objective and four subjective ending conditions proposed by [27]. Third, we ran through five iterations until all IoT platform providers from the database were classified and the ending conditions were met (figure 1). In the first iteration, we opted for a conceptual-to-empirical iteration and integrated the taxonomy dimensions identified during the literature review (table 1). For the second, third, fourth, and fifth iterations, we conducted empirical-to-conceptual cycles, leading to successful classification of all IoT platform providers in the data sample. To reduce the complexity of these cycles, we classified the providers according to their company size i.e., multinational companies (MCN), small and medium companies (SME), startups and project settings (e.g., open source projects). At this point, all providers from the database were classified, seven further dimensions were added (i.e., customer type, industry focus, operational level, core technology, operation mode, pricing model, continuous revenues) and all objective and subjective ending conditions were fulfilled, thereby ending the iterations.

3.4 Phase 4

In the fourth phase, we conducted a cluster analysis to empirically derive IoT platform business model archetypes from the taxonomy. Based on the work of [28], a two-step approach delivers the best results. They indicate that using Ward's method to identify the optimal number of clusters represents the first step. Therefore, the clusters must be further specified using an iterative partitioning procedure. In accordance with this recommendation, we first applied the Ward's clustering algorithm (all analyses were conducted in SPSS, version 22). The similarities between the two organizations (objects) were calculated by the number of identical characteristics along the taxonomy dimensions and measured as square Euclidean distance, which is suitable for binary data. These results indicate that five or nine cluster solutions would be most useful for a k-means in depth analyses.

We then applied the k-means method to the five and nine cluster solutions. In both cases, the algorithm ran through five iterations until no significant improvements were achieved in the last iteration. Subsequently, we evaluated the resulting clusters manually for their explanatory power. A closer look reveals that the five cluster solution is largely similar to the seven core capabilities of the IoT platforms neglecting some important differences between underlying business logics. The nine clusters, on the other hand, were more fine-grained and reflected some essential differences, particularly regarding value capture. In short, the nine-cluster solution has more explanatory power because it provides important distinguishing characteristics. Therefore, we will only explain the results of the nine-cluster solution in this article.

4 Results

4.1 Taxonomy

The resulting taxonomy contains 15 dimensions, each with two to seven different characteristics (table 2). Each of the 195 platform providers from the database are represented by exactly one characteristic per dimension. The taxonomy contains the most important dimensions according to which IoT platform providers differ i.e., business model components that are identical for all providers are not listed here [27].

Value Proposition

Core capabilities: What are the core capabilities offered for developers by the IoT platform? First of all, the capabilities of IoT platforms greatly vary. Where some platforms focus on things by providing operating systems to develop embedded software for smart devices, other platforms enable the connectivity of things by offering telecommunication hardware such as sim cards and the infrastructure to manage the connections (e.g., Ciscos Jasper Platform). Some others focus on the management of devices by allowing remote firmware updates and configuration management of the connected devices (e.g., Echelon). Another group of platforms offer cloud storage options for the incoming device data or running code (e.g., GroveStreams) whereas others mainly provide analytic capabilities including means to extract insights from the data (e.g., SenseIoT) or capabilities to develop and deploy IoT applications through a developer environment (e.g., Ayla). A last type of IoT platform provides multiple capabilities such as, application development environments and cloud storage to provide an all-in-one IoT platform approach (e.g., ThingWorx).

Table 2. Taxonomy of IoT platform business models

VISOR	Dimension	Characteristics of the dimensions						
Value proposition	Core capabilities	Embedded device operation	Connectivity enablement	Device management	Device data storage	Analytics	Application development	Multiple capabilities
	Device support	Selected 3rd party devices	Exclusively provider´s devices	Selected 3rd party devices and provider devices	Any device, if provider's standards are used			
	Customer type	Consumer	Business	Business and consumer				
	Industry focus	Single-industry platform	Cross-industry platform					
Interface	Platform integration	In enterprise system	In web services	In multiple diverse systems	No integration opportunities			
	Application sales channel	Marketplace functionality	External marketplace necessary					
	Platform openness	Fully proprietary	Hardware proprietary	Software proprietary	Open source			
Service platform	Operational level	Operated on device	Operated on cloud	Operated on device and cloud				
	Core technology	Telecommunications	Sensors and microcontroller	Cloud technologies	Other technologies			
Organizing model	Partner system	Open partner system	Proprietary partner system	No partner system				
	Degree of support	Non personal technical support	Personal technical support	Personal technical and business support				
	Operation mode	Operated by platform provider	Operation by 3rd party possible					
Revenue Model	Pricing model	Developer projects are free and enterprise projects are priced	Developer and enterprise projects are priced	Free for use				
	Transaction based revenues	Per connected device	Per API call	Traffic based	Combination of multiple sources	Per request	Free for use	
	Continuous revenues	Time based (monthly / yearly) minimum fees	Pure pay as you grow (no continuous fees)					

Device support: What are the devices supported by the platform? Some platform providers allow selected third-parties to connect their devices to the platform, whereas others exclusively support platform provider devices. Several platforms support all types of connected devices when the platform standards (libraries) are implemented on the device and the predefined API is used.

Customer type: What is the customer type of the platform? Even though a majority of the platform providers are focused on business customer types (e.g., General Electrics), some providers also address the platform (e.g., IFTTT) to the consumer segment. *Industry focus:* What is the operating industry of the platform? Some platforms are designed to operate in a single specific industry, e.g., home or manufacturing, whereas others are designed as a cross-industry platform.

Interface

Platform integration: Does the platform offer integration functionality for other systems? While some platforms offer a strong functionality to connect to enterprise software like ERP, CRM or SCM systems, other platforms provide integration frameworks to enable rapid connectivity to web services, e.g., like AWS.

Application sales channel: Does the platform offer a marketplace to publish IoT applications? Some providers offer a marketplace (e.g., Libelium) to support developers in publishing applications. On other platforms, the developers rely on external marketplaces e.g., Arktik (Samsung), Pega 7 (Pegasystems).

Platform openness: To what extent is the platform open for modifications? Some platform providers follow the open source approach, i.e., the platform code can be modified and is open to run on any third-party device (e.g., KAA IoT platform). In contrast, there are platforms that are hardware proprietary and run exclusively on specific devices (e.g., Siemens Mindsphere), or software proprietary, which means that they can operate on any device with no access to the platform code. Some platforms are also fully proprietary, meaning that developers cannot run the platform on third party devices or change the platform code (e.g., Ayla IoT platform).

Service Platform

Operational level: On what level is the platform operated? IoT platforms run on different operating levels. Some run on a central cloud and manage millions of devices (e.g., Grovestreams), whereas others can be operated on edge devices such as a local gateway or small single-board computers (e.g., Phao, Fosstrack). Furthermore, some other platforms have a flexible structure and can run on both levels (e.g., Nimbits).

Core technology: What core technology does the provider supply along with their platform? The core technology may include sensors and microcontrollers (e.g., Siemens, General Electrics), telecommunications technologies (e.g., Cisco, Huawei, Vodafone), cloud infrastructure (e.g., Amazon, Google), or other technologies.

Organizing Model

Partner system: How is the access to the partner system organized? Some providers operate a proprietary partner system which does not allow the direct participation of complementary providers. Other providers offer direct registration for the partner program, taking into account complementary competencies and services. Some providers do not operate a partner system at all.

Degree of support: What kind of support does the platform provider offer? Some platform providers offer non-personal technical support e.g., via documentations or an online forum, whereas others run personal technical support teams. Some providers give additional business support, for example, in finding an IoT business strategy.

Operation mode: Who operates the platform? Some providers exclusively operate the platform on their own, whereas others also allow operation by third parties, for instance, on the customer's infrastructure or an external cloud service provider.

Revenue Model

Pricing model: What pricing models does the platform provider follow? Platform providers mostly distinguish between restricted developer projects (e.g., including

limited connected devices) and professional enterprise projects (e.g., large scale projects). Some platform providers do not charge the developer projects and solely price enterprise projects whereas others charge both kinds of projects. A few providers offer their platforms free of charge (e.g., KAA, Node Red, Paho).
Transaction based revenues: What are the transaction based revenues received by the platform provider? Platform providers make revenues on different kinds of transactions: e.g., per connected device, per API call, or per generated traffic. Other providers combine different sources or calculate the price per specific request.
Continuous revenues: Is the platform provider receiving continuous revenues? Some platform providers charge a fixed fee regardless or in addition to the actual transactions (hybrid pricing). Others follow a pure “pay as you grow” approach and charge no minimum fees.

4.2 Archetypes

The analysis identified nine clusters, each cluster covers between 15 to 32 IoT platforms of the 195 data samples. The extracted clusters have different centers along the dimensions and characteristics of the IoT platform business model taxonomy. As the taxonomy is mutually exclusive and collectively exhaustive, the results can be presented as percentages (table 3). For instance, in the sixth Cluster, 94% of the platforms offer application development as a core capability for developers, whereas 6% of the IoT platforms provide analytic capabilities for incoming device data. In the following, we present the characteristics that differentiate the nine clusters.

Cluster 1 - Embedded Device Operation Platform: Platforms in the first cluster provide capabilities to develop and operate embedded software on devices. As devices are equipped with more sensors, data processing power and increased connectivity, developers must manage the emerging complexity. Embedded device platforms support developers in overcoming the underlying device complexity, as the platforms ensure the integration of diverse hardware components and provide the functionality to develop add-on applications on the basis of the platform’s core. RIOT OS, is a typical example for such an embedded device platform and represents an operating system (similar to Linux or Windows for the personal computer) that can be operated on any connected device. Other embedded device platforms in the first cluster are Amazon (FreeRots), Microsoft (Windows IoT) and Googles (Android Things).

Cluster 2 - Device Connectivity Enablement Platform: Platforms from the second cluster enable and manage the connectivity of devices via telecommunication technologies (e.g., sim cards, router and gateways). The Cisco Jasper platform, with 40 million connected devices is probably the largest provider of this cluster. The platform ensures that all devices used by developers are connected in a reliable and secure manner. It also provides the software environment for efficiently managing the connectivity to all devices that are equipped with a cisco global sim. The platform is addressed fully to large scale business customers and operates in several industries (e.g., smart security). Other representatives of the second cluster are Telekom (Cloud of Things), or Ericsson (Device Connection Platform).

Cluster 3 – Device Management Platform: The third cluster describes platform providers who ensure that the connected devices are working accurately on a device level by running patches, updates for software as well as changing embedded configurations. For example, the Helix Device Cloud operated by the technology provider Wind River (Intel subsidiary). The platform supports developers in device monitoring, bidirectional file transfer, remote access as well as in the detection and diagnosis of device problems mainly for devices included in Wind Rivers partner system (e.g., ARM, Intel). The Helix Device Cloud is focused on business customers who develop solutions in several industries (transportation, building automation, street lighting) and provides integration opportunities in existing ERP and CRM systems. Other well-known industrial platform providers in the third cluster are Sierra Wireless (AirVantage), Siemens (MindSphere) and Relayr.

Cluster 4 - Device Data Storage Platform: The fourth cluster consists of platform providers who deliver scalable cloud infrastructure for the storage of device data and operating codes to developers. The SenseIoT platform, launched in 2015 and a spinoff of Sense OS, offers a storage platform where any device can connect to, by using the SenseIoT API and store streamed device data. As storage platforms are designed in a universal manner, they address consumers as well as business customer and run in multiple industries (no restrictions). SenseIoT is a fully proprietary platform that is operated on a cloud level. SenseIoT, as it goes for the fourth cluster, is a pure "pay as you grow" provider. Other prominent providers offering cloud storage platforms include Sales Force (IoT Cloud) or Pentaho IoT.

Cluster 5 - Device Data Analytics Platform: The fifth cluster includes platforms that perform various device data analyses based on millions of incoming device data sets, from data clustering to machine learning up to predictive analysis. GroveStreams, a platform on which IoT analyses run, offers a range of tools for analyzing and visualizing IoT device data and operates a typical business model from this cluster. Similar to the fourth cluster (cloud storage platforms), Grovestreams allows any device to be connected to the platform, and allows developers to use it in multiple industries. Other representatives of the fifth cluster are Uptake, DataRPM or Cloudera.

Cluster 6 - Application Development Platform: The sixth cluster describes platform providers offering IoT application enablement through a software development kit to enable rapid development and deployment of IoT applications. These platforms offer a set of tools that incorporate language-independent support for programming-in-the-large tasks. The platform provider, Ayla Networks, operates the Ayla IoT platform that is a typical representative of the sixth cluster. The platform contains a variety of APIs that can be used to rapidly develop iOS and Android applications based on incoming device data. Ayla operates the platform and charges enterprises as well as developer projects per customer specific request. Other platforms belonging to the sixth cluster are DeviceHub, Arktik (Samsung) and Temboo.

Cluster 7 - Application Development, Market Place Platforms: The seventh cluster covers platforms that offer application development capabilities (similar to the sixth cluster) for developers. A typical business model from this cluster is operated by IFTTT. The platform is a free cloud service that allows application building on the basis of the if-this-then-that logic (IFTTT). For instance, if a user possesses an intelligent lighting system, it can be linked to the users' location. If the user leaves the house, the light switches itself off. And when the user gets back home, the lights turn on by themselves. IFTTTs support all devices if the platforms standards are used and is one of the few platform providers who offer, a market place functionality that allows developers to publish their if-this-then-that applications. Other examples of the seventh cluster are the more business focused IoT marketplaces of Libelium, PTC or Telus, where hardware components are also offered in addition to applications.

Table 3. Cross cluster analysis

Dimensions	**Characteristics**	**Cluster** 1	2	3	4	5	6	7	8	9
Number of platform providers per cluster		19	15	25	25	27	*32*	*22*	*15*	15
Core capabilities	Multiple capabilities	0%	0%	0%	0%	0%	0%	0%	13%	100%
	Application development	0%	0%	4%	0%	0%	94%	100%	87%	0%
	Analytics	0%	0%	8%	0%	100%	0%	0%	0%	0%
	Device data storage	0%	0%	0%	100%	0%	6%	0%	0%	0%
	Device management	0%	0%	88%	0%	0%	0%	0%	0%	0%
	Connectivity enablement	0%	100%	0%	0%	0%	0%	0%	0%	0%
	Embedded device operation	100%	0%	0%	0%	0%	0%	0%	0%	0%
Device support	Selected 3rd party devices	0%	0%	0%	8%	7%	0%	0%	0%	0%
	Exclusively provider's devices	0%	0%	4%	0%	0%	9%	0%	0%	0%
	Selected 3rd party devices and provider devices	5%	67%	76%	8%	19%	0%	0%	20%	13%
	Any device, if provider's standards are used	95%	33%	20%	84%	74%	91%	100%	80%	87%
Customer type	Consumer	0%	0%	0%	0%	0%	0%	9%	0%	0%
	Business	100%	100%	88%	20%	100%	94%	14%	0%	87%
	Business and consumer	0%	0%	12%	80%	0%	6%	77%	100%	13%
Industry focus	Single-industry platform	0%	0%	12%	4%	0%	9%	0%	13%	0%
	Cross-industry platform	100%	100%	88%	96%	100%	91%	100%	87%	100%
Platform integration	In enterprise systems	0%	87%	80%	76%	0%	0%	0%	0%	20%
	In web services	16%	0%	0%	8%	15%	97%	100%	33%	0%
	In multiple diverse systems	0%	0%	20%	16%	85%	0%	0%	0%	80%
	No integration opportunities	84%	13%	0%	0%	0%	3%	0%	67%	0%
Application sales channel	Marketplace functionality	0%	0%	0%	0%	0%	0%	100%	0%	13%
	External marketplace necessary	100%	100%	100%	100%	100%	100%	0%	100%	87%
Platform openness	Fully proprietary	0%	100%	88%	100%	15%	94%	86%	0%	80%
	Hardware proprietary	0%	0%	4%	0%	85%	0%	9%	0%	20%
	Software proprietary	37%	0%	4%	0%	0%	6%	5%	0%	0%
	Open source	63%	0%	4%	0%	0%	0%	0%	100%	0%
Operational level	Operated on device	100%	0%	0%	0%	0%	0%	0%	0%	0%
	Operated on cloud	0%	100%	12%	80%	93%	100%	95%	33%	100%
	Operated on device and cloud	0%	0%	88%	20%	7%	0%	5%	67%	0%
Core technology	Telecommunications	0%	93%	8%	12%	7%	9%	0%	0%	20%
	Sensors and microcontrollers	5%	0%	76%	4%	7%	3%	0%	0%	27%
	Cloud technologies	16%	0%	8%	72%	74%	13%	5%	0%	27%
	Other technologies	79%	7%	8%	12%	11%	75%	95%	100%	27%
Partner system	Open partner system	95%	20%	88%	96%	11%	97%	86%	0%	93%
	Proprietary partner system	5%	80%	4%	0%	89%	0%	9%	0%	7%
	No partner system	0%	0%	8%	4%	0%	3%	5%	100%	0%
Degree of support	Non-personal technical support	100%	0%	4%	16%	96%	78%	0%	93%	0%
	Personal technical support	0%	100%	96%	76%	0%	6%	9%	7%	0%
	Persona technical and business support	0%	0%	0%	8%	4%	16%	91%	0%	100%
Operation mode	Operated by platform provider	0%	100%	100%	100%	100%	97%	95%	13%	93%
	Operation by 3rd party possible	100%	0%	0%	0%	0%	3%	5%	87%	7%
Pricing model	Dev. projects free and enterprise projects priced	0%	0%	0%	0%	0%	100%	77%	0%	7%
	Dev. and enterprise projects priced	0%	100%	96%	100%	100%	0%	23%	7%	93%
	Free for use	100%	0%	4%	0%	0%	0%	0%	93%	0%
Transaction based revenues	Per connected device	0%	13%	4%	4%	4%	6%	5%	0%	0%
	Per API call	0%	0%	4%	16%	7%	3%	9%	0%	0%
	Traffic based	0%	0%	0%	60%	11%	6%	5%	0%	0%
	Combination of multiple sources	0%	20%	68%	8%	78%	0%	0%	0%	0%
	Per request	0%	67%	12%	12%	0%	72%	0%	0%	93%
	Free for use	100%	0%	12%	0%	0%	13%	91%	100%	7%
Continuous revenues	Time based (monthly/ yearly) minimum fees	0%	87%	96%	20%	4%	0%	9%	13%	100%
	Pure pay as you grow (no continues fees)	100%	13%	4%	80%	96%	100%	91%	87%	0%

Cluster 8: Application Development, Open Source Platforms: The eighth cluster also describes IoT application development platforms. However, in contrast to the sixth cluster, providers from this cluster typically offer an open source freeware IoT platform for developers. A well-known example and a typical platform of this cluster is Nimbits, launched in 2014. Nimbits supports developers in application building by providing an easy to use web server that connects things to each other. Based on the open source approach, Nimbits can be operated on any device (from a single-board computer to a cloud service) and hence, by third parties. Other freeware open source application development platforms in this cluster include KAA Project and DeviceHive.

Cluster 9 - Multi Capability Platforms: Platforms of the ninth cluster comprise multiple capabilities, such as an environment to rapidly build frontend applications (cluster 6), to manage devices (cluster 3) as well as to analyze incoming device data (cluster 5). A typical example of this all in one platform approach is ThingWorx that was acquired by PTC in 2013. ThingWorx supports all devices that use the provided code library and covers several businesses mainly in the industrial area. ThingWorx provides strong integration capabilities in existing enterprise systems (e.g., of SAP and web services as AWS). Furthermore, platforms from this cluster typically offer personal technical as well as business support, e.g., ThingWorx employs an "IoT university" to support developers in monetizing their IoT applications. Other platforms belonging to the ninth cluster are OceanConnect (Huawei), Bosch IoT Suite or Microsoft Azure IoT.

5 Discussion

5.1 Implications for Research and Practice

Our study is the first to focus on IoT platforms as an essential building block of IoT by using a business model lens and empirically deriving archetypes of an IoT platform business model. Thus, our taxonomy presents an overview of the phenomenon by combining the most important dimensions mentioned in existing research e.g., device support, application sales channel, or the degree of platform support [2, 21] with newly identified dimensions that are essential to structure and compare the business logics of IoT platforms (e.g., customer focus, core technology, and pricing model). By abstracting beyond the business model of individual platform providers, our taxonomy helps to identify different types of IoT business models based on their dimensions and characteristics. In addition, the taxonomy as well as the archetypes allow for a quick understanding of important differences, and assist researchers in anchoring their contributions more precisely within these types. The iterative taxonomy development process devised by [27] allows other researchers to extend the presented taxonomy when new IoT business models arise in the future. In short, our taxonomy and archetypes provide new insights into the design options of IoT platform business models and help systematize and synthesize previous fragmented research at the IoT.

Our archetypes provide findings regarding the issue of how different hardware (e.g., telco and device manufactures) and software firms (e.g., cloud providers) strategize in a platform environment [8]. As business model archetypes can conceptually be viewed as linkages between the business strategies of the platform providers and their

operational strategy implementation [12], our results contribute to the emerging body of work regarding digital business strategies [29]. For instance, hardware manufacturers (cluster 2, 9) and telecom operators (cluster 3, 9) have been implementing a bottom up strategy by positioning their IoT platforms on top of their existing device offerings. By launching their platforms, those firms extend their existing offering (telco hardware, sensors, and microcontrollers) and add another opportunity for digital value creation while simultaneously shifting away from a hardware-centric business model. In contrast, software based companies have been following a top down approach by providing platforms for analytics, storage and application enablement or combining a bottom up (embedded device platform) as well as bottom down (multi capability platform) approach. For instance, Microsoft and Amazon have been providing an embedded device platform (cluster 1) as well as multi capability platforms (cluster 9) and can therefore provide the capabilities to operate, connect and build applications on top of existing devices, without manufacturing any. Third, there are startups that mainly operate IoT platforms within single product strategy. Their IoT platforms typically emerge with a system centric approach, often focusing on a top down strategy that incorporates capabilities to develop applications or managing devices based on code library's which is a facilitator for building up scalable, extensible, heterogeneous systems (e.g., Thing Worx) often supported by acquisition or funding of incumbent firms (e.g., PTC). However, as the interconnection of people and things progresses and numerous companies exploit the economic potentials, our archetypes provide a fruitful starting point of how different actors strategize in the IoT platform ecosystem.
Our research also provides important contributions for managerial practice. First, our database, taxonomy and archetypes offer managers a comprehensive overview of the fast moving IoT platform market. Furthermore, the taxonomy and the archetypes facilitate a quick understanding of the most common business model configurations of IoT platform providers. Second, our taxonomy serves as a concrete tool for business model innovation as it allows the necessary abstraction needed to identify unoccupied business models, as reflected by combinations of characteristics currently not offered by competing firms. We acknowledge that the taxonomy serves as a tool for simulating creativity but is not the key to reveal the perfect business model.

5.2 Limitations and Future Research

Our study is not free of limitations. First, [27] describe a taxonomy as never truly perfect, but in the best case useful. We argue that our developed taxonomy is useful in dissecting the business models of IoT platform providers. It also helps researchers and practitioners to identify differences within a business model configuration. Additionally, it is necessary to note that there is no perfect number of clusters and they are at best useful for a specific purpose. Second, the results of our research are limited to the business model configuration of the IoT platform providers at the time of data collection. As the IoT platform market was almost non-existent a few years ago and changes quickly, the database and the classification of individual providers will therefore become obsolete soon. However, the taxonomy and the archetypes of the IoT platform business models reflect the constituent elements of these instances and make them relevant for a longer period. Third, as we build our taxonomy within the empirical to conceptual cycles on the database, our study is inherently limited to the IoT platform providers contained in our sample and the respective information given at the provider's website. Though our database is quite large, following a consistent understanding of the research object and containing leading platform providers, not every platform provider has been listed in our dataset, nor can the reliability of the stated information be ensured. Fourth, given absence of a standard description of IoT platforms, the provider's information is often distributed across the entire website, using an inconsistent terminology and forcing the researchers to derive the complete information. However, as the study progressed, researchers became more experienced in handling this information and deriving complete datasets.

Given the current dynamic of the IoT field, a need for deeper and future-oriented investigations in the field is evident. Thus, the archetypes and corresponding characteristics can serve as starting points for studies on the transformation of the identified business model archetypes (e.g., extending core capabilities). In addition, when linking the identified business model archetypes to financial figures, they support the analyses of IoT platform providers with regard to aspects such as the probability of long-term success or the funding received, or the acquisition patterns of individual archetypes vis-à-vis others. Furthermore, the consideration of platform operating costs (not covered by the VISOR framework) in a future analysis will result in a more comprehensive understanding of the identified archetypes. However, based on the generality of our approach, following the aim to fit different types of IoT platforms, we had to compromise on the level of granularity. Future research can elaborate on precise taxonomies, by focusing on specific archetypes of IoT platform providers.

References

1. Gartner: Competitive Landscape of IoT Platform Vendors, https://www.gartner.com/doc/3730917/competitive-landscape-iot-platform-vendors (Accessed 27.10.2017)
2. Mineraud, J., Mazhelis, O., Su, X., Tarkoma, S.: A gap analysis of Internet-of-Things platforms. Computer Communications 89-90, 5–16 (2016)
3. Amazon: Miovision Case Study – Amazon IoT, https://aws.amazon.com/solutions/case-studies/miovision/ (Accessed 12.09.2018)
4. IoT Analytics: IoT Platform Comparison: How the 450 providers stack up, https://iot-analytics.com/iot-platform-comparison-how-providers-stack-up/ (Accessed 25.08.2018)
5. Wortmann, F., Flüchter, K.: Internet of Things. Bus Inf Syst Eng 57, 221–224 (2015)
6. Borgia, E.: The Internet of Things vision. Key features, applications and open issues. Computer Communications 54, 1–31 (2014)
7. Püschel, L., Schlott, H., Röglinger, M.: What's in a Smart Thing? Development of a Multi-layer Taxonomy. In: Proceedings of the 37th International Conference on Information Systems, pp. 1–19, Dublin (2016)
8. Reuver, M. de, Sørensen, C., Basole, R.C.: The digital platform: a research agenda. J Inf Technol 89, 1–12 (2017)
9. Osterwalder, A., Pigneur, Y., Clark, T.: Business model generation. John Wiley & Sons, Hoboken, New Jersey (2010)

10. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
11. Foss, N.J., Saebi, T.: Fifteen Years of Research on Business Model Innovation. Journal of Management 43, 200–227 (2016)
12. Zott, C., Amit, R., Massa, L.: The Business Model. Recent Developments and Future Research. Journal of Management 37, 1019–1042 (2011)
13. Fichman, R.G., Dos Santos, B.L., Zheng, Z.(E.): Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum. MIS Quarterly 38, 329–353 (2014)
14. Al-Debei, M.M., Avison, D.: Developing a unified framework of the business model concept. Eur J Inf Syst 19, 359–376 (2010)
15. El Sawy, O.A., Pereira, F.: VISOR. A Unified Framework for Business Modeling in the Evolving Digital Space. Springer, Berlin, Heidelberg (2013)
16. Weill, P., Woerner, S.L.: Optimizing Your Digital Business Model. MIT Sloan Management Review 54, 71–78 (2013)
17. Remane, G., Nickerson, Robert, C., Hanelt, A., Tesch, J., Kolbe, Lutz, M.: A Taxonomy of Carsharing Business Models. In: Proceedings of the 37th International Conference on Information Systems, pp. 1–19, Dublin (2016)
18. Peters, C., Blohm, I., Leimeister, J.M.: Anatomy of Successful Business Models for Complex Services: Insights from the Telemedicine Field. Journal of Management Information Systems 32, 75–104 (2015)
19. Webster, J., Watson, R.T.: Analyzing the Past to prepare for the Future: Writing a Literature Review. MIS Quarterly 26, xiii–xxiii (2002)
20. Bandyopadhyay, S., Sengupta, M., Maiti, S., Dutta, S.: Role of Middleware for Internet of Things. International Journal of Computer Science & Engineering Survey 2, 94–105 (2011)
21. Díaz, M., Martín, C., Rubio, B.: State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing. Journal of Network and Computer Applications 67, 99–117 (2016)
22. Emeakaroha, V.C., Cafferkey, N., Healy, P., Morrison, J.P.: A Cloud-Based IoT Data Gathering and Processing Platform. In: Proceedings of the 3rd International Conference on Future Internet of Things and Cloud, pp. 50–57, Rome (2015)
23. Lempert, S., Pflaum, A.: Towards a Reference Architecture for an Integration Platform for Diverse Smart Object Technologies. In: Proceedings of the 6th conference on Mobile and Ubiquitous Information Systems, pp. 53–66, Kaiserslautern (2011)
24. Nakhuva, B., Champaneria, T.: Study of Various Internet of Things Platforms. International Journal of Computer Science & Engineering Survey 6, 61–74 (2015)
25. Ray, P.P.: A survey of IoT cloud platforms. Future Computing and Informatics Journal 1, 35–46 (2016)
26. Sruthi, M., Kavitha, B. R.: A survey on IoT Platforms. International Journal of Scientific Research and Modern Education 1, 1-29 (2016)
27. Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. Eur J Inf Syst 22, 336–359 (2013)
28. Punj, G., Stewart, D.W.: Cluster Analysis in Marketing Research. Review and Suggestions for Application. Journal of Marketing Research 20, 134–148 (1983)
29. Bharadwaj, A., El Sawy, O., Pavlou, P., Venkatraman, N.: Digital Business Strategy. Toward a Next Generation of Insights. MIS Quarterly 37, 471–482 (2013)

Revitalizing established Industrial Companies: State of the Art and Success Principles of Digital Corporate Incubators

Matthias Hille[1], Matthias Lederer[2], Dominik Forster[1]

[1] Institute of Information Systems, University of Erlangen-Nürnberg, Nürnberg, Germany
{matthias.hille, dominik.forster}@fau.de

[2] ISM International School of Management, Department Economics & Quantitative Methods
matthias.lederer@ism.de

Abstract.
In order to compete with radical digital ideas, more and more large companies are founding corporate incubators. These small and agile companies are designed to accelerate disruptive innovation and transfer it to the parent company. However, it is obvious that the design and work in these newly established units are often efficient and the return of ideas cannot be effectively realized. This paper has collected a list of 131, partly non-public, incubators and evaluated them according to classic design parameters for companies (e.g., location, industry, etc.). In addition, success factors for digital corporate incubators (such as the design of the innovation process) were collected for deriving success principles.

Keywords.
Corporate incubators, innovation management, start-up management, innovation lab, new business development

1 Introduction

A lot of "traditional" and established industries are facing quick changes with possibilities of digitalization [1]. It is accepted in literature and business practice, that disruptive innovation is hardly carried out in long-existing and large-sized companies, which often rely on fundamental approaches within their Research and Development (R&D) departments [2]. This leads to a common trend: With the help of special organizational units, internal entrepreneurs or external startups, established companies try to force digital innovation [3].

Many established companies (e.g. Siemens, Lufthansa, Bayer) have set up incubators, which operate mostly independent from other business units [4]. The term "corporate/business incubator" is frequently used for incubators, which are connected to a company [3]. Research on corporate incubators is a relatively new area and

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

incubators in companies have not been categorized [5]. The effectiveness and results of business incubators are controversially discussed, and success factors are not defined or even known [6]. Often parent companies interfere too much [7] or give too little budget as well as wrong resources [8]. Successful examples show that even supposedly small factors such as the location or the legal form can have a major impact on success. The outcome of the innovation and venture creation process has high uncertainty and needs flexible management [9]. Incubators and accelerators are, especially in business and technology-driven countries, highly needed [10].

The motivation for this work-in-progress contribution is to answer the following research questions (RQ):

RQ1: What types of corporate incubators do currently exist worldwide?

RQ2: What are the environmental preconditions of successful incubators?

2 Methodology

A combination of approaches was used following the mixed methods by [11] to identify patterns in hidden research field data.

1. **Identification**: In the first step, the authors created a list of currently implemented corporate incubators using a general desk research. Available sources such as the Deutscher Aktienindex (DAX), published case studies, documented interviews, as well as book chapters, were used to collect and further investigate incubators. Another approach was interviewing managers of corporate companies at international fairs, and in personal meetings about their current activities. Fragmented lists of incubators (e.g., list by [8]) were consolidated. The generated list seems to be the currently largest dataset on corporate incubators available and will be part of the poster presentation.
2. **Coding**: All incubators identified were further analyzed by internet research. Metadata (e.g. name, parent company, etc.), as well as relevant parameters (e.g. preconditions as budget, location, number of employees), were documented. These factors result from a literature review on fragmented success factors of incubators in scientific databases.
3. **Analysis**: Since the extension of the dataset is an ongoing research of the authors, the dataset as well as the interviews conducted with many incubators, a qualitative content analysis based on [12] was performed so far.

3 Preliminary results

Some statistics on the dataset and derived success will shortly be introduced. The four different categories of incubators shown in Figure 1 are based on the work from [10] and [5] and was adapted with the results from performed interviews and statistics. The, in literature, used category names are combined with the knowledge of several interviews to *Independent*, *Corporate-integrated*, *Communal/State* and *Academical incubators*.

3.1 Relative Distribution

Based on the list of 131 collected incubators, statistical analyses were possible. For several areas of interest, clear industry trends are visible.

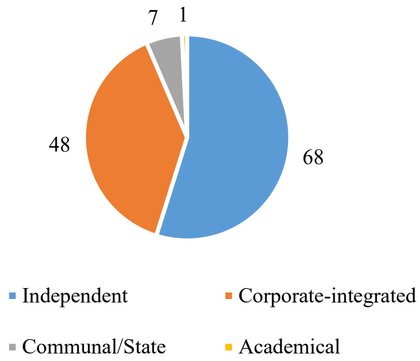


Figure 1. Incubator types

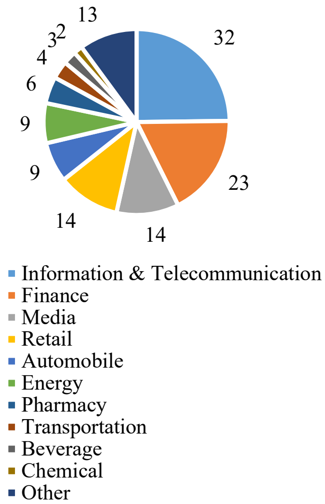


Figure 2. Industry fields of incubators

Figure 1 visualizes that most of the listed incubators are independently founded as separate company. Some are part of the formal parent organization (corporate-integrated). Only a small amount of analyzed incubators are communal/state (7) or academical (1) incubators. The high number of independent is an expected result as the research for incubators focused on corporate/business incubators. This is a reasonable result as communal/state driven incubators and academical incubators are not profit-oriented; they support local economics or regional institutes and cooperation with companies. Most incubators identified are in Germany and the USA. Within industries large companies establish incubators whose market offers are extensively digitized (Figure 2). Furthermore, incubators are found in those industries where significant changes are expected by radical technological progress.

3.2 Environmental preconditions of successful incubators

Besides the statistical analysis, four main design areas were identified during qualitative interviews based on [12], which can have a positive influence on corporate incubators, when they receive attention during the strategic setup. The interviews were performed with structured questionnaires and the answers were analyzed in the three steps *Clustering*, *Extraction* and *Aggregation* based on [13].

1. **Strategic focus**: Incubators need a clear strategic orientation in order to be able to prioritize and decide on a project. It was mentioned by managers of

incubators that existed more than 2 years have a lot of tasks from the existing parent company instead of focusing on entirely new business opportunities.

2. **Intervention phase**: To successfully focus on few but relevant and promising projects and business models, it is necessary to have criteria how possible projects are evaluated. Similarly, to principle 1 the most responsible managers of incubators mentioned in interviews that they had to establish an evaluation framework to validate and prioritize new ideas.
3. **Preferred exit-path**: For several interviewees, the spin-off exit path was especially the favourited exit-path for disruptive innovations in the 10% area. Especially for incubators from industrial companies, the exit path was mentioned in direct interviews as important.
4. **Corporate integration**: The transfer of projects into existing corporate structures needs proper configuration as problems can occur during the transfer. A success factor is to define the transfer into corporate departments with all stakeholders in the corporate structure. Some interviewees mentioned that the transfer must be defined during the initial configuration.

4 Implications for further research

This work in progress provides a comprehensive dataset on digital corporate incubators worldwide. The analysis of the success factors is in an initial stage and needs more work to be based on a broader information basis. As shown in 3.2 the invention phase is, of course, important for corporate incubators, but the exit path and the corporate integration is also of interest. The results and methods from this research in progress can serve as a general foundation for discussions with other participants of the research area at the conference. The generated list of incubators and examples that were evaluated more in detail can be used for further research with adaptions based on the discussion and remarks from the conference. The researched incubators have a strategical focus on corporate integration and independency as this is one key for medium- and long-term success. Regarding a low amount of white papers and industry studies more innovation labs are switching focus from corporate projects to company building.

References

1. Mertens, P., Barbian, D.: Digitalisierung und Industrie 4.0–Trend mit modischer Überhöhung?. Informatik-Spektrum 39, 301-309 (2016)
2. Branstad, A.: A study of management tasks and stakeholders in a hybrid corporate incubator. European journal of innovation management, 13, 294–312 (2010)
3. Becker, B., Gassmann, O.: Gaining leverage effects from knowledge modes within corporate incubators. R&d Management 36, 1–16 (2006)
4. von Zedtwitz, M.: Classification and management of incubators: aligning strategic objectives and competitive scope for new business facilitation. International Journal of Entrepreneurship and Innovation Management, 3, 176–196 (2003)

5. Gassmann, O., Becker, B.: Towards a resource-based view of corporate incubators. International journal of innovation management 10, 19–45 (2006)
6. Kanbach, D. K., Stubner, S.: Corporate Accelerators As Recent Form Of Startup Engagement: The What, The Why, And The How. Journal of Applied Business Research 32, 1761–1776 (2016).
7. Schuh, G., Vogt, F., Lau, F., Bickendorf, P.: Concept of Innovation Transfer from Corporate Incubators. In: 2017 Portland International Conference on Management of Engineering and Technology, pp. 1-11. PICMET, Portland (2017).
8. Heinemann, F.: Corporate Accelerators: A Study on Prevalence, Sponsorship, and Strategy. Technical report, Massachusetts Institute of Technology (2015)
9. Stegmann, P., Zec, M., Matthes, F.: Assessing the Potential Value of Software-Support for the Venture Creation Process, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1417-1428 (2017)
10. Spath, D. (Ed.): Mehr Innovationen für Deutschland: wie Inkubatoren akademische Hightech-Ausgründungen besser fördern können. Springer-Verlag. (2012)
11. Johnson, R. B., Onwuegbuzie, A. J.: Mixed methods research: A research paradigm whose time has come. Educational researcher 33, 14-26 (2004)
12. Mayring, P.: Qualitative Inhaltsanalyse. VS, Wiesbaden (2010)
13. Mayring, P., & Fenzl, T.: Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), Handbuch Methoden der empirischen Sozialforschung (pp. 543–556). Wiesbaden: Springer Fachmedien Wiesbaden. (2014)

When 1+1 is Greater than 2: Concurrence of Additional Digital and Established Business Models within Companies

Jonas Toutaoui

TU Darmstadt, Institute of Information Systems & E-Services, Darmstadt, Germany
toutaoui@ise.tu-darmstadt.de

Abstract. Many established companies currently face digitalization challenges. Part of their answer is often the creation of new digital business models based on new technologies, which do not necessarily replace the existing business model but act as additional source of economic value. Having two business models in parallel in the same company creates the opportunity of synergies between these business models. However, knowledge about the interaction between digital and non-digital business models remains scarce. This study, currently in progress, contributes to research by filling this gap and identifying synergies already taken by companies of various industries, leading to a better understanding of digital business models within established companies and their interaction with existing, traditional business models. For practitioners this study provides insights on how to benefit from parallel business models within a company and thus how to better face digitalization.

Keywords: Digital business model, business model, synergy, digitalization

1 Introduction

BMW created DriveNow as IT-enabled business model [1] offering car-sharing. However, BMW didn't terminate its existing business model of manufacturing cars and therefore has two different business models in the same company, one traditional and one digital. In fact, as new digital business models often go hand in hand with companies' answer to digitalization and as many companies have not yet faced digitalization challenges [2], even more new digital business models and co-existence between the new digital business model and the established business model can be expected [3].

While research attention in the past focused on conflicts, synergies and mechanisms on how to manage two non-digital business models in the same company [4, 5], these business models differ mostly through low-cost vs. premium products [6]. However, digital business model relate to value creation, delivery and capture through new digital technologies [7, 8] and the potential differentiation to traditional business models is large. So far researchers shed light on the interaction between digital and non-digital business models primarily through the differentiation of channels (offline vs. online) in

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

publishing and retailing [9, 10]. This leads to our research question: “What potential synergies exist between a new additional digital business model and the established business model?”

Through multiple interviews with managers from six companies from various industries, this paper identifies synergies between digital and established business models based on business model representation and the synergy theories of complementarities and resource relatedness [11-13].

The study at hand contributes to research by providing a better understanding of IT-enabled digital business models within established companies facing digitalization, answering the research call of Veit et al. [14]. Lastly, it supports practitioners to better manage and ensure the success of a new digital business model.

2 Theoretical Background

2.1 Digitalization

Digitalization can be described as the manifold sociotechnical phenomena and processes of adopting and using new digital technologies in broader individual, organizational, and societal contexts [15]. As of today this digitalization is a major challenge for many established companies from different sectors [2]. Hence, this topic is of importance for practitioners but is also recognized by IS scholars as it has reached top IS journals [16]. The focus is mostly on digital business strategies companies can employ to handle the digitalization [17, 18], often affecting changes in value creation with the use of new technologies leading to new digital business models.[19]

2.2 Business Model Representation

Many definitions for business models exist [20], the commonly accepted definition by Al-Debei et al.[14, 21] is: “The business model is an abstract representation […] of all core interrelated architectural, co-operational, and financial arrangements designed and developed by an organization [and] all core products and/or services the organization offers” (p.372). Furthermore, a digital business model is based on new technologies to create, deliver and capture value [7] e.g., video-on-demand services [14], departing from the differentiation of online vs. offline channels [10].

The business model representation from Steiniger et al. [22] (adapted from [7]) is used as he showed that its suitability to compare non-digital and digital business models without being too abstract. The nine components, grouped in four categories, of the business model representation are finally:

1. Product: Value Proposition
2. Customer Interface: Target customer, Distribution channel, Customer relationship
3. Infrastructure Management: Capability, Value configuration, Partnership
4. Financial aspects: Cost structure, Revenue Model

2.3 Synergies

The concept of synergy primarily comes out of the strategy and economics research [13]. The two types of synergies are super-additive value synergy and sub-additive cost synergy.

Super-additive value synergy is based on a complementary set of resources and can be described following the economic theory of complementarities as "doing more of one thing increases the returns to do more of another" [12]. Sub-additive cost synergy refers to the use of common resources across units and is therefore based on resource relatedness [11].

Similar to Radszuwill and Fridgen [23], the definitions of synergies can be adapted to the business model context with (a) and (b) being two business models:

- *Super-additive business model value synergy*: When the sharing of business model components between two or more business models leads to increased value compared to conducting the business models individually (Value(a+b) > Value(a)+Value(b)).
- *Sub-additive business model cost synergy*: When the sharing of business model components between two or more business model leads to lower costs compared to conducting the business models individually (Costs(a+b) < Costs(a)+Costs(b)).

3 Research methodology

This study follows the interpretive research principles of Klein and Myers [24] for qualitative, idiographic research using a multi-case approach [25-27]. We selected companies from various industries and size to allow for replicability. Only established companies having an additional digital business model for more than two years were chosen to ensure that they can sufficiently inform the research [28]. In each of the six cases two semi-structured interviews, based on an interview guideline, were conducted, on site or on phone, with two managers to discover different viewpoints and remove biases of interview partners [29, 30]. Additional documentation, publicly available or shared by interview partners served as secondary data source. Coding is done following the business model representation components as first order coding elements [31, 32]. Table 1 summarizes the cases and positions of interview partners.

Table 1. Selected cases and interview partners

Case ID	**Sector**	**Size** (number of employees)	**Interview partner**
1	High-tech	10,000 – 50,000	Head of Digital innovation unit
			Project manager within IT
2	Pharma-ceutical	> 50,000	Manager within digital innovation unit
			Team leader within IT
3	Automotive	> 50,000	Manager within digital innovation unit
			Team leader within IT

Case ID	Sector	**Size** (number of employees)	Interview partner
4	Utilities	10,000 – 50,000	Manager within digital innovation unit
			Team leader within IT
5	Logistics	< 10,000	Manager within digital innovation unit
			Chief Information Officer
6	Automotive	> 50,000	Head of Digital Innovation Unit
			Team leader within IT

4 First Results

Our initial research reveals potential synergies along several business model representation components and we focus in this section on case 1 only, to demonstrate first results. In case 1 the company built up a subscription-based platform, offering image recognition and analytics of images and data collected via medical technology of different manufacturers as new digital business model. Sales and maintenance of high-tech, e.g. medical technology underlie its traditional business model.

Super-additive business model value synergy: Through the new digital business model, the *value proposition* of the traditional business model is enhanced as existing products can be directly connected to the platform, thus offering higher value to similar costs. Also, in terms of *capabilities* the traditional business model benefits as knowledge about new capabilities of the digital business model such as cloud development or agile way of working are shared.

Sub-additive business model cost synergy: The development of the new digital business model is accelerated thanks to solid and reliable *distribution channels* within the traditional business model. In fact, existing channels (e.g., via industry fairs) were used to not only promote the traditional business model but also the new digital business model.

5 Discussion and Conclusion

This paper at hand can be seen as an initial examination of synergies between additional digital business models and established business models within companies. Despite being limited by the number of cases, first results indicate that the traditional business model benefits from super-additive synergies (e.g., complementary products) whereas the new digital business model is supported by sub-additive synergies (e.g., resource relatedness). Further analysis will create a more comprehensive overview of synergies, along all business model components, contributing to research and practice by a better understanding of new additional digital business models, grounded on synergy theory, and of companies' answer towards digitalization. Future research might focus on an in-depth quantitative analysis of the impact of business model synergies on the performance of a new digital business model.

References

1. Remane, G., Hanelt, A., Hildebrandt, B., Kolbe, L.M.: Discovering new digital business model types: a study of technology start-ups from the mobility sector. Proceedings of the 20th Pacific Asia Conference on Information Systems, Taiwan (2016)
2. Remane, G., Hanelt, A., Wiesböck, F., Kolbe, L.M.: Digital Maturity in Traditional Industries - An Exploratory Analysis. Proceedings of the Twenty-Fifth European Conference on Information Systems (ECIS), Portugal (2017)
3. Hanelt, A., Piccinini, E., Wayne Gregory, R.W., Hildebrandt, B., Kolbe, L.M.: Digital Transformation of Primarily Physical Industries – Exploring the Impact of Digital Trends on Business Models of Automobile Manufacturers. Proceedings of the 12th International Conference on Wirtschaftsinformatik, pp. 1313-1327, Germany (2015)
4. Markides, C., Charitou, C.D.: Competing with Dual Business Models: A Contingency Approach. The Academy of Management Executive 18, 22-36 (2004)
5. Smith, W.K., Binns, A., Tushman, M.L.: Complex Business Models: Managing Strategic Paradoxes Simultaneously. Long Range Planning 43, 448-461 (2010)
6. Porter, M.E.: What Is Strategy? Harvard Business Review 74, 61-78 (1996)
7. Osterwalder, A., Pigneur, Y.: An e-Business Model Ontology for Modeling e-Business. Proceedings of the 15th Bled Electronic Commerce Conference, (2002)
8. Osterwalder, A., Pigneur, Y.: Designing Business Models and Similar Strategic Objects: The Contribution of IS. Journal of the Association for Information Systems 14, 237-244 (2013)
9. Hoßbach, N.: Dual Business Models: Going beyond Spatial Separation. In: Twenty-first American Conference on Information Systems (AMCIS). (Year)
10. Wiener, M., Hoßbach, N., Saunders, C.: Omnichannel businesses in the publishing and retailing industries: Synergies and tensions between coexisting online and offline business models. Decision Support Systems 109, 15-26 (2018)
11. Hill, C.W.L., Hoskisson, R.E.: Strategy and Structure in the Multiproduct Firm. The Academy of Management Review 12, 331-341 (1987)
12. Milgrom, P., Roberts, J.: Complementarities and fit Strategy, structure, and organizational change in manufacturing. Journal of Accounting and Economics 19 179 -208 (1995)
13. Tanriverdi, H.: Performance Effects of Information Technology Synergies in Multibusiness Firms. MIS Quarterly 30, 57-77 (2006)
14. Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J.M., Loos, P., Spann, M.: Business Models. Business & Information Systems Engineering 6, 45-53 (2014)
15. Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N., Ahlemann, F.: Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. Business & Information Systems Engineering 59, 301-308 (2017)
16. Gerster, D.: Digital Transformation and IT: Current State of Research. In: PACIS. (Year)
17. Bharadwaj, A., El Sawy, O.A., Pavlou, P.A., Venkatraman, N.: Digital Business Strategy: Toward a Next Generation of Insights. MIS Quarterly 37, 471-482 (2013)
18. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for Formulating a Digital Transformation Strategy. MIS Quarterly Executive 15, (2016)
19. Matt, C., Hess, T., Benlian, A.: Digital Transformation Strategies. Business & Information Systems Engineering 57, 339-343 (2015)
20. Wirtz, B.W., Pistoia, A., Ullrich, S., Göttel, V.: Business Models: Origin, Development and Future Research Perspectives. Long Range Planning 49, 36-54 (2016)

21. Al-Debei, M.M., El-Haddadeh, R., Avison, D.: Defining the Business Model in the New World of Digital Business. Proceedings of the Fourteenth Americas Conference on Information Systems, Toronto (2008)
22. Steininger, D.M., Hoelz, J., Veit, D.: Comparing Traditional and Electronic Business Models of the Music Industry: A Content Analytical Approach. Proceedings of the ConLife Academic Conference, (2012)
23. Radszuwill, S., Fridgen, G.: Forging a Double-Edged Sword: Resource Synergies and Dependencies in Complex IT Project Portfolios. Proceedings of the Thirty Eighth International Conference on Information Systems (ICIS) South Korea (2017)
24. Klein, H., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly 23, 67-94 (1999)
25. Benbasat, I., Goldstein, D., Mead, M.: The Case Research Strategy in Information Systems. MIS Quarterly September, (1987)
26. Yin, R.K.: Case Study Research: Design and Methods (2003)
27. Walsham, G.: Interpretive case studies in IS research: nature and method. European Journal of Information Systems 74-81 (1995)
28. Bilgeri, D., Wortmann, F., Fleisch, E.: How Digital Transformation Affects Large Manufacturing Companies' Organization. Proceedings of the Thirty Eighth International Conference on Information Systems, South Korea (2017)
29. Keutel, M., Michalik, B., Richter, J.: Towards mindful case study research in IS: a critical analysis of the past ten years. European Journal of Information Systems 23, 256-272 (2017)
30. Benlian, A., Haffke, I.: Does mutuality matter? Examining the bilateral nature and effects of CEO–CIO mutual understanding. The Journal of Strategic Information Systems 25, 104-126 (2016)
31. Remane, G., Hanelt, A., Nickerson, R.C., Kolbe, L.M.: Discovering digital business models in traditional industries. Journal of Business Strategy 38, 41-51 (2017)
32. Klostermeier, R., Haag, S., Benlian, A.: Digitale Zwillinge – Eine explorative Fallstudie zur Untersuchung von Geschäftsmodellen. HMD Praxis der Wirtschaftsinformatik (2018)

Special Track 01:

Student Track

Track Chairs:

Prof. Dr. Jan vom Brocke
Universität Liechtenstein

Dr. Bernd Schenk
Universität Liechtenstein

Dr. Markus Weinmann
Universität Liechtenstein

Investigating Personalized Price Discrimination of Textile-, Electronics- and General Stores in German Online Retail

Tsagana Badmaeva[1], Joschka Andreas Hüllmann[1]

[1] WWU Münster, Interorganizational Systems Group, Münster, Germany
{t_badm01,huellmann}@uni-muenster.de

Abstract. Developers of pricing strategies in e-commerce businesses see a wide range of opportunities for deploying online price discrimination techniques given their ability to track consumers' online identity and behavior. In theory, an increasing use of personal data enables organizations to show every single consumer their own personalized price, which is determined by the consumer's characteristics, e.g. age, gender, surfing history, or location. This paper aims to explore the existence of online price discrimination activities within the German e-commerce market using a three-method approach. First, inquiring the online retailers via email and investigating their public documents; second, surveying students; and third, using a software crawler to simulate surfing activity. Our results do not provide any evidence of individualized price discrimination, which, we argue, is due to economic and political reasons, not technical reasons.

Keywords: online price discrimination, tracking, privacy, e-commerce, personalization.

1 Introduction

The German e-commerce market increases steadily as people are more comfortable and willing to purchase goods online[1]. At the same time, the trend for personalization in online retail is growing [1, 2]. We see prospects of e-commerce businesses to collect and evaluate personal data such as age, gender, location or device used in order to offer specific prices for each customer – called *personalized price discrimination*. However, the public is wary of this development and raises concerns about privacy, tracking and consumer protection. Institutions such as the OECD [3], the German Federal Ministry of Justice [4], consumer protection initiatives [5, 6], newspapers [7, 8], and the citizens question what effects data and algorithms have on consumers. For a debate on price discrimination, it is vital to uncover discriminatory practices in order to make interac-

[1] https://www.statista.com/statistics/453490/e-commerce-retail-revenue-share-germany/ (last access 2018-10-12).

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

tions with online retailers transparent. Using personal data for setting prices is a sensitive issue, in particular in Germany [9], which could require political or legislative measures in the future [10].

Previous studies [6–8] explored personalized pricing practices of internationally operating e-commerce retailers, ranging from hotel-booking sites to general online retailers, by using software scrapers. However, the studies do not focus on the German market, while existing German studies prefer a manual approach to collect data and have small sample size [9–11]. We lack comprehensive information on the current status of personalized pricing in German online retail. To close this gap, we suggest triangulating results from manual data collection and automatic scraping.

In this paper, we explore the prevalence of personalized pricing activities performed by textile-, electronics- and general stores in the German online retail market. We selected the eleven biggest online retailers by revenue and investigated the presence of price discrimination practices in their shops by combining surveys, an automated crawler and a discourse analysis. Our study contributes an overview of the tracking ecosystem and reports on the existence of personalized pricing activities in German online retail. It provides survey items and a software crawler to identify such mechanisms. Our research does not detect any price discrimination practices for the selected online retail shops, which, we argue, results from economic and political reasons.

2 Background

2.1 Price Discrimination

Effective pricing strategies are key to increase demand, sales and profit [16]. One pricing strategy is price discrimination (or differential pricing). In general, price discrimination is defined as a variation of price cost ratio across units or groups of buyers [17]. In our context, we define price discrimination as follows: the producer sets varying prices, which includes discounts, for an identical product or service for different consumers based on the consumer's characteristics, the time or location of purchase, the amount of purchase, or other relevant information [14], [16–18]. The goal is that the consumer pays the highest amount that they are willing to pay, i.e. the marginal willingness to pay (WTP) [21]. Price discrimination requires information about consumers and the producer’s ability to estimate the consumer’s marginal WTP as accurately as possible [19]. In an ideal information environment, the producer can set the price to the consumer’s marginal WTP to extract the maximum consumer surplus and maximize his profit. Depending on the available information, three degrees of price discrimination are distinguished [17, 22]:

Third degree price discrimination describes different prices for segments of consumers and is the most common form of price discrimination. It requires sociodemographic information about the consumer [17]. Examples are student or senior citizen discounts or regional price variations. **Second degree price discrimination** is based on the *amount* or *quality* of a given product or service, usually including consumers’ self-selection of *amount* or *quality*, to mitigate the difficulty of distinguishing consumer types. Examples for *amount* are volume discounts and examples for *quality* are airfares, namely business and economy tariffs. **First degree price discrimination** requires the most information as each person gets a personalized price based on their characteristics [17]. With detailed information about the consumer, the producer can determine the maximum WTP and capture the maximum consumer surplus. Collecting enough information to implement first degree price discrimination is difficult. Nonetheless, large-scale data collection is a first step to enable first degree price discrimination. For example, Shiller [23] demonstrated a profitable first degree price discrimination model based on 5000 trackable consumer characteristics using behavioral and sociodemographic traits. Historically, this type of price discrimination is known from the used cars or insurance market or from the bazaar in one-to-one transactions. Now, it is possible to implement and automate it for all consumers online.

The degrees are not mutually exclusive [19]. Personalized price discrimination is based on the segments of third degree price discrimination, with the ultimate goal to achieve a first degree price discrimination. Alternatively, we can explain personalized price discrimination as an extreme case of third degree price discrimination with segments of size one.

2.2 Conditions and Implementation of Price Discrimination

Certain conditions must be met to implement price discrimination. A firm (1) needs some market power, or consumers buy elsewhere, (2) it needs to control the sale of the product or service to set the price and (3) the consumers' willingness to pay must differ from one another [17]. Sufficient information about consumers to distinguish them into segments must be available [24, 25]. The segments need to be actionable, substantial, accessible, measurable, profitable and stable and are distinguished by behavioral, psychographic, sociodemographic and geographic criteria (for more information see [23–25] [2]). Behavioral criteria include used channels, brand loyalty, buy volume, and previous purchases. Psychographic criteria include lifestyle, social identity and personality, for example, attitude towards risk, expectations of quality of products. Sociodemographic information includes gender, age, children, job, education, income. Geographical criteria are distinguished into macro (country, city) and micro (district, street). Based on data for each customer, segments can be inferred from quantitative statistical analysis such as cluster analysis or machine learning algorithms [26, 27].

The data is sourced internally or acquired externally. Internal data includes customer master data from CRM systems, i.e. data that a consumer enters on the website, for example, name, address, city, etc. Other data is based on the actions that customers perform, i.e. which articles they view, click, purchase, rate and review. Besides, the data includes technical information such as IP address, device, user-agent, or operating system. Prevalent shop software solutions offer the functionality to collect this data, and to use it for consumer segmentation as well as price discrimination. In addition to analyzing internal data, shop solutions offer integrations with external data sources.

[2] Market segmentation is a concept that goes beyond price discrimination.

External data sources originate either from third party analytics services or trackers[3], which capture users' online activities on the web [28] and may provide extended insights [29, 30], or from offline datasets, e.g. voter records, or vehicle owner data. There are multiple services available, for example, TruSignal, Adobe Customer Experience Cloud, Google Analytics 360, Bluekai, Lotame, and more. However, the market of these services is highly concentrated around a few big players, who accumulate data, which makes it valuable to acquire, but also sensitive to privacy issues [29–31].

2.3 Consumer Perspective

Growing consumers' awareness of aggressive tracking and the amount of gathered data lead to a discussion on the privacy and ethics of tracking and personalized prices. This is exacerbated by the fact, that the big players accumulate most of the data [32, 33], and the extent of data collection and sharing is concealed from consumers [30].

Since online retailers are not transparent about their pricing mechanisms, it is unknown if personalized pricing takes place at all, and what type of data significantly changes the segmentation of consumers and the resulting prices [15]. The growing data collection and the lack of transparency on what consequences particular information has, lead to privacy issues and may unsettle consumers [34]. There is an information asymmetry between the online retailers and the consumers [31]. Consumers can only believe the privacy statements, which are mandated since the EU-GDPR and given by online shops, as there are no (easy) neutral ways to check for personalized price discrimination [35]. For a debate on personalized price discrimination in Germany, it is crucial to provide empirical means to investigate the current extent of personalized price discrimination in online retail. Therefore, we contribute such means to assess the extent of personalized price discrimination, empirically, and report on our results.

3 Related Work

Previous studies have found little to none price discrimination in online retail. Despite checking 200 U.S. shops, Mikians et al. [12] found only few price differences caused by a combination of system, sociodemographic and behavioral criteria. Hannak et al. [11] explored U.S. shops and the tourism industry using Amazon Mechanical Turk and simulated user profiles. They found different prices on nine of sixteen e-commerce websites due to changes in system and behavioral criteria. Vissers et al. [36] and Hupperich et al. [13] investigated the tourism industry, both using simulated browser fingerprints and user profiles. Vissers et al. did not find any price discrimination, while Hupperich et al. find differences based on geographic criteria. For both studies, other criteria did not lead to any price changes. Schleusener and Hosell [4] and Kraemer et al. [14] focused on the German market using a manual approach. The former found

[3] Due to space limitations, we cut the technical description of third-party trackers and the tracking ecosystem. Instead, an explanation is provided in [38, 39].

price discrimination in the tourism industry, with system criteria being relevant. Interviews with professionals corroborated their results. Kraemer et al. [14] found price discrimination based on geographic and system criteria.

The findings of previous studies are inconclusive, as some results indicate price discrimination in the tourism industry, while other studies find no price discrimination at all. Furthermore, only two studies focus on the German online retail market, leaving it underexplored and requiring further research.

4 Research Design

To investigate the existence of price discrimination in the online retail market, we select the eleven biggest German online e-commerce retailers according to their annual revenue. They are the most relevant shops from a consumer rights perspective, i.e. more revenue means more involved consumers, and we choose textile, electronics and general stores, because they are less explored compared to the tourism industry. First, we analyze the discourse on price discrimination using public information on the selected shops, shop software vendors, tracking providers followed by contacting them. Second, we build a survey based on the found information, which captures personal user information. Test persons filled out the survey and use their personal devices to check prices in each shop, manually. Third, we use an automated software crawler, which simulates user activity and user profiles, to check the prices in each shop.

Table 1. Selected Online Shops (Revenue in Million EUR in 2015)[4]

Shop	*Revenue*	*Sector*	*Shop*	*Revenue*	*Sector*
amazon.de	7.790,60	General	tchibo.de	450,00	General
otto.de	2.300,00	General	conrad.de	433,20	General
zalando.de	1.031,80	Textile	alternate.de	376,70	General
notebooksbilliger.de	610,90	Electronic	hm.de	344,60	Textile
cyberport.de	491,30	Electronic	esprit.de	342,00	Textile
bonprix.de	484,70	Textile			

First, for the discourse analysis, we investigate the most prevalent trackers among the top 20 worldwide [33, 37] and observe the involved actors as well as their activities. We evaluate public documents on their websites in order to understand what, how and why they track the information of consumers in their services. Besides the trackers, we look at the three biggest standard shop software by market share in 2017 (Shopify, WooCommerce, Magento)[5]. We evaluate their websites for functionality related to tracking, consumer segmentation and price discrimination; and what consumer criteria the software makes use of. We also contact them directly, and assess available third

4 http://www.statista-research.com/wp-content/uploads/Infografik_Top-100-Onlineshops_D.pdf (last access 2018-10-09).

5 https://www.statista.com/statistics/710207/worldwide-e-commerce-platforms-market-share (last access 2018-10-09).
The market share has changed since we collected the data.

party addons for each shop software. According to the EU-GDPR, e-commerce companies have to disclose how they obtain, process and use personal citizens' information as well as the purposes of doing so. This implies that companies are to disclose whether they perform any kind of price discrimination activities using personal data of customers. Using this opportunity, we assess the privacy notices of each online shop and contact the data protection departments of each online shop. The question that we forwarded to the data protection departments is:

"*Due to the EU-GDPR, we would like to inquire, whether and in which way personal data influences the determination of visible prices, including potential discounts or vouchers (also via mail or email), in your online shop.*" (Translated from the original German statement).

Second, the three checked shop software packages provide the functionality to implement personalized price discrimination, but there is no information if any of the shops use this feature. Only the number of reviews hints that the features may be used, so we use the criteria collected from the discourse analysis to develop a survey. The survey consists of two parts. The first part is a list of 20 hyperlinks to products of the selected shops. The second part asks for behavioral and sociodemographic criteria as well as system attributes as seen in Table 3. We distributed the survey to four different student groups in Germany depicted in Table 2. The students checked the prices of the 20 selected products with their own devices, browser settings, and accounts and filled out the survey accordingly. Some students used proxies and VPNs, but the most common types of connection were University WiFi and mobile data. Due to logistical issues, we were unable to record the type of connection.

Table 2. Samples

Round	*Group*	*Size*
1st round	International students	12
	PhD students	14
2nd round	Bachelor students	13
3rd round	Bachelor students	22

Table 3. Survey Items

Survey Items	
Age	Manufacturer
Gender	Hobby
Device Type	Previous Visits
Operating System	Previous Purchases
Language	Social Media Logins
Apps	

Within the groups, at least 10 different participants checked each product's price. All checks were conducted in one session at the same time, to rule out prices changes over time. Between the groups, different products and shops were tested to increase the breadth of the study. The participants recorded all prices in EUR, which we later analyzed.

Third, due to the logistics of conducting a survey, we develop an automated software to triangulate our findings based on [12, 13]. The goal of the software is to simulate ordinary user activity for seven days, to build five online user profile and accumulate cookies. After seven days, it checks the prices using the five profiles.

Table 4. Crawler Profiles

Profile	*Keywords*
Profile1	Frankfurt Football
Profile2	Tennis Dortmund
Profile3	Horseriding München
Profile4	Cooking News
Profile5	Gucci Rolex

Table 5. Actions for Profile Simulation

Website	*Engagement Actions*
twitter	click and like
reddit	click, like and subscribe
facebook	click, view and like
google	click, scroll and wait
bild	click, scroll and wait
youtube	click, view and like

Each user profile is generated from two keywords (Table 4) that the crawler uses to engage on search, news and social media sites (Table 5), and is assigned a distinct proxy to simulate geographic location. The software is implemented in Python using Selenium with the Firefox driver and the source code is available on GitHub.

5 Findings

Discourse Results. The tracking ecosystem sees actors with varying primary activities, from data management platforms to marketing solutions, who track consumers' information. We found that most of them have dedicated services that perform tracking. The tracked data is distinguished into sociodemographic, behavioral and system criteria, sourced internally and externally. The purposes of tracking by the actors overlap. The criteria in Table 6 depicts what data the clients of the actors can acquire, not what data the firms behind the services collect and analyze.

Table 6. Players in Tracking and Targeted Analytics. Legend: Google Analytics 360[G], Facebook Analytics[F], Amazon Pinpoint[P], Adobe Target[T], Optimizely[O], SAP Hybris Marketing Cloud[H].

Main Service	Data Management Platform[G], Advertising[G,F,O], Marketing Solution[T,H]
Data Sources	Internal[G,F,P,T,O,H], Across Tools[G,P], External[G,F,P,T,O,H], CSV[G]
Sociodemographic	Age[G,F,P], Gender[G,F], Location[G,F,P,T,O], Name[H], Address[H], Other[H]
Behavioral	Surf History[G,P,O], Lifestyle[G,P,H], Interests[G,P,H], # of Visits[P], Time of Visit[T,O], Referrer[O], Purchases[T,H], Payments[T,H], Clicked ads [G], On-site Activity[O,H]
System	IP address[G,F,T,O], Device[G,P,O], Manufacturer[P,O], Browser[F,T,O], Operating System[F,P,T,O], Screen size[T], Language[T,O]
Purpose	Segmentation[G,P,O,H], Ad Effectiveness[F,T,O,H], Engagement[T,O,H], Site Optimization[T,O], Loyalty Management[T,H]

We reached out to the three online shop vendors to inquire whether their software supports personalized price discrimination. Although it is not a core functionality of Shopify and WooCommerce, it is feasible through addons. For Shopify, multiple addons are available, for example, "Customer-Specific Pricing", "Storakle", or "Segment

Builder", while for WooCommerce, we only found "WP Statistics". Magento has advanced customer segmentation built-in, which can be enhanced by the addon "Prices per Customer". Magento also allows the integration of external data, for example, the import of offline purchase data.

We found that all of the observed online retailers authored a privacy declaration, in which they depict data use for targeted advertising, statistics and server operations, but not for pricing purposes. After inquiring about the use of price discrimination via email, the majority of the contacted online retailers (7) denied usage of discrimination activities towards customers based on their personal data, while Otto.de, Bonprix.de, hm.de and Amazon.de did not answer the mail at all. Tchibo.de and Esprit.de state that they select consumers for discounts and gifts based on personal data and previous purchases, i.e. they engage in indirect price discrimination based on ex post discounts.

Survey Results. Due to space limitations, we do not report all descriptive statistics in detail. We only report them for selected items and only the top three values: n=61, age (M=24.84; SD=4.5), gender[6] (64% male; 18% female), device type (34% Windows; 31% Android; 13% iOS), and language (62% German; 26% English; 12% other). Most selected hobbies are travelling (46%), technology (41%) and sports (20%) and most visited shops include Amazon.de (43%), Zalando.de (28%) and Notebooksbilliger.de (25%). Top installed apps are YouTube (56%), Facebook (46%) and Instagram (38%). The first round with 26 international and PhD students showed a few insignificant prices differences. For four distinct users, prices differed by 9.00 EUR, 2.00 EUR, 0.05 EUR and 0.09 EUR on Otto.de, Zalando.de and Cyberport.de (last two). For the biggest difference, 9.00 EUR and 2.00 EUR, we compared the survey characteristics of the users with differing prices. Although, both of them were using mobile devices, the model and language of those devices were distinct from the other test persons. The second round of tests, in which 13 bachelor students participated, showed two differences in prices at 0.05 EUR and 0.60 EUR for Tchibo.de and Alternate.de. The last round with 22 bachelor students found only one instance of price differentiation by 0.10 EUR on Amazon.de. Overall, we tested 93 prices of products. Due to insufficient cases of price differentiation – presumably caused by transcription errors – the results do not substantiate the existence of systematic online price discrimination in the observed online retailers.

Crawler Results. Utilizing the crawler, we scraped the prices of 865 products with each of the five profiles, totaling 4325 checks in 27 sessions. We crawled all prices within the sessions at the same time, while the different sessions took place throughout one month. The shops Notebooksbilliger.de, Hm.de and Conrad.de are missing in the results due to their bot protection. They recognized our crawler and prevented the page load and scraping of prices. To add another electronics store, we crawled Mediamarkt.de instead. Apart from software errors, when the website layout changed and the scraper failed to extract the correct price, we did not find a single difference in prices in the entire data set for any of the selected products and shops.

[6] Eleven participants did not fill out their gender.

6 Discussion

Our results are consistent with previous studies and show that personalized price discrimination is not widely used in online shops. Nevertheless, as evidenced by the practices of the tourism industry [4, 11, 13], the implementation of personalized price discrimination is generally feasible. The first key condition to implement it, is sufficient information on the consumers to distinguish them into segments [19, 24, 25]. Based on the current data collection means, enough information on the consumers is available to derive such segments and to implement third degree price discrimination [17, 22]. Schleusener and Hosell [4] hypothesize a reason for the lack of personalized price discrimination may be the lack of technical expertise by German online retailers. However, our results show that all standard software packages provide the means to collect internal data and offer integrations with external data providers. Furthermore, if the shops have the expertise to develop a custom shop software, then they have the expertise to implement consumer segmentation and price discrimination. Hence, we argue that the lack of expertise is not the reason for the non-deployment of price discrimination practices.

Instead of technical obstacles, the reasons may be economical. The data providers do not disclose any prices on their websites, but acquiring external data may be costly compared to the value it provides, so acquiring external data may not be profitable for online shops. Yet, this does not explain the lack of price discrimination using internal data. As stated in the second condition for price discrimination [17], the retailer needs market power, or consumers will switch to other shops, if the same products are offered elsewhere. This is exacerbated by readily available online search. Another reason may be that employing personalized price discrimination can hurt the brand [4], because changing prices can lower the perceived price fairness of consumers [16].

Online retailers acquiring and exchanging customer data with third parties (including outside of EU legislation), without knowledge by the consumers may bring about dissatisfaction [10]. German citizens, compared to other countries and cultures, are sensitive to privacy issues, in particular since the EU-GDPR is effective [9]. The threat of legal action by consumer protection initiatives, and potential repercussions due to mistreatment or theft of data, may bar online retailers from implementing personalized price discrimination. It may be too risky to lose the trust of consumers and their brand value over the comparatively small profits, delivered by personalized price discrimination practices.

We conclude that the reasons for the lack of price discrimination in Germany are not technical, rather it must be for economic and political reasons. The digitalization, paired with intransparency of data collection and usage, especially by international players, brings rapid changes and means that price discrimination can start to occur any time. Thus, it is imperative to monitor potential price discrimination continuously. In a bigger project, our software can be improved with better monitoring abilities to provide ongoing assessment of personalized price discrimination.

We also received anecdotal references from colleagues and consultants, which we were not able to corroborate. Contrary to our findings, three retail and e-commerce consultancies claimed that their clients are engaging in online price discrimination. We assume that either their clients have not rolled out any price discrimination yet, or they are in pre-testing, or they are not part of our selected shops, or our methods did not detect them.

7 Conclusion

In previous studies, German shops were not widely tested using surveys and automated scraping, and the ecosystem with its players was not systematically explored. Hence, our study makes a valuable contribution, as we investigate the prevalence of personalized price discrimination practices in the German e-commerce market through triangulation. By using three different methods, we reach a holistic picture of the ecosystem, its players and their use and exchange of data for the purpose of price discrimination. Contrary to our expectations, the results of our study do not substantiate that online retailers in Germany perform online price discrimination practices. In particular, the discourse analysis shows that the technical possibilities for price discrimination are widely available. Yet, both the survey and the crawler results show no evidence of price discrimination by the tested online shops. As discussed, we argue that this is due to political and economic reasons, not because of technical reasons.

The discourse analysis is limited to the official documentation, references and responses of the players. The survey and crawler are restricted by the sample size. Other limitations include the colocation of the survey participants and the usage of University Wi-Fi, since one of the characteristics that was proven to alter prices within consumers is the geographic location. Some participants made errors while transcribing prices. Several shops detected the crawler as a bot. We used free proxies, which might be blacklisted and negatively affect the price scraping, and the crawler directly accessed the product pages and did not traverse the shop websites. Nevertheless, our results match those observed in earlier studies. Future iterations of this experiment should be conducted with larger sample size and an improved crawler. The reasons for price discrimination can be further explored by conducting expert interviews.

Various institutions such as the OECD [3], the German Federal Ministry of Justice and Consumer Protection [4], consumer protection initiatives [5, 6], newspapers [7, 8], and most importantly, the German citizens and consumers recognized price discrimination as a major concern. Personalized price discrimination is a prime example of the algorithmization and digitalization of society and economy, and what effects data and algorithms may have on its people. Due to the digitalization and rapid changes of the online ecosystem, we encourage a continuous assessment of price discrimination practices in the future.

8 Acknowledgements

We thank Felix Murrenhoff, who programmed the original version of the crawler and collected data.

References

1. Odlyzko, A.: Privacy, Economics, and Price Discrimination on the Internet. Ssrn. (2003).
2. Klein, S., Loebbecke, C.: Emerging pricing strategies on the web: lessons from the airline industry. Electron. Mark. 13, 46–58 (2003).
3. OECD: Algorithms and Collusion: Competition policy in the digital age. (2017).
4. Schleusener, M., Hosell, S.: Expertise zum Thema “Personalisierte Preisdifferenzierung im Online-Handel.” 1–26 (2015).
5. Verbraucherzentrale des Bundesverbands: Personalisierte Preise | Diskussionspapier des vzbv | 21. September 2016. , Berlin (2016).
6. Dautzenberg, K., Gaßmann, C., Groß, B., Müller, F., Neukamp, D., Bodenstein, U.: Dynamische Preisdifferenzierung im Deutschen Online-handel. , Brandenburg (2018).
7. Theile, G.: Preise im Internet - Wie man die Tricks der Online-Händler zum Sparen nutzen kann, http://www.faz.net/aktuell/finanzen/meine-finanzen/geld-ausgeben/spar-tipps-diese-tricks-helfen-beim-online-shopping-15724854.html, (2018).
8. Beenen, J., Wilke, F.: Einkaufen im Netz - So rasant ändern Online-Händler die Preise, https://www.sueddeutsche.de/wirtschaft/einkaufen-im-netz-so-rasant-aendern-online-haendler-die-preise-1.4080347, (2018).
9. EU publications: Consumer conditions scoreboard. (2017).
10. Zuiderveen Borgesius, F., Poort, J.: Online Price Discrimination and EU Data Privacy Law. J. Consum. Policy. 40, 347–366 (2017).
11. Hannak, A., Soeller, G., Lazer, D., Mislove, A., Wilson, C.: Measuring Price Discrimination and Steering on E-commerce Web Sites. Proc. 2014 Conf. Internet Meas. Conf. - IMC ’14. 305–318 (2014).
12. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Detecting price and search discrimination on the internet. In: Proceedings of the 11th ACM Workshop on Hot Topics in Networks - HotNets-XI. pp. 79–84. ACM Press, New York, New York, USA (2012).
13. Hupperich, T., Tatang, D., Bochum, R., Wilkop, N., Bochum, R., Holz, T.: An Empirical Study on Online Price Differentiation. In: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy. pp. 76–83 (2018).
14. Krämer, A., Kalka, R., Ziehe, N.: Personalisiertes und dynamisches Pricing aus Einzelhandels- und Verbrauchersicht. Mark. Rev. St. Gall. (2016).
15. Genth, S., Schleusener, M., Kenning, P., Pohst, M., Remmel, J., Weber, B., Gier, N., Schmidt-Kessel, M.: Dynamische Preissetzung — Wer profitiert? Wirtschaftsdienst. 96, 863–882 (2016).
16. Lii, Y., Sy, E.: Computers in Human Behavior Internet differential pricing: Effects on consumer price perception, emotions, and behavioral responses. Comput. Human Behav. 25, 770–777 (2009).
17. Carroll, K., Coates, D.: Teaching Price Discrimination: Some Clarification. South. Econ. J. 66, 466–480 (1999).
18. Hoffman, K.D., Turley, L.W., Kelley, S.W.: Pricing retail services. 55, 1015–1023 (2002).
19. Varian, H.R.: Differential Pricing and Efficiency. First Mondey J. 1, 1–10 (1996).
20. Armstrong, G.: Marketing: an introduction. Pearson Education (2009).
21. Stirling, D.A.: Information rules. (2003).
22. Pigou, A.C.: The economics of welfare. Macnillam, London (1920).
23. Shiller, B.R.: First-Degree Price Discrimination Using Big Data. Brand. Univ. 1–36 (2014).
24. Fassnacht, M.: Preisdifferenzierung. In: Hermann, A. and Diller, H. (eds.) Handbuch Preispolitik. pp. 483–502. Gabler Verlag, Wiesbaden (2003).

25. Meffert, H., Burmann, C., Kirchgeorg, M.: Marketing - Grundlagen marktorientierter Unternehmensführung: Konzepte-Instrumente-Praxisbeispiele. Springer Fachmedien Wiesbaden, Wiesbaden (2015).
26. Simon, H.: Preismanagement. Gabler Verlag, Wiesbaden (1992).
27. Wedel, M., Kamakura, W.A.: Market Segmentation: Conceptual and methodological foundations. Springer Science & Business Media (2012).
28. Mayer, J.R., Mitchell, J.C.: Third-Party Web Tracking: Policy and Technology. In: 2012 IEEE Symposium on Security and Privacy. pp. 413–427. IEEE (2012).
29. Puglisi, S., Rebollo-Monedero, D., Forné, J.: On Web User Tracking: How Third-Party Http Requests Track Users' Browsing Patterns for Personalised Advertising. arXiv. (2016).
30. West, S.M.: Data Capitalism: Redefining the Logics of Surveillance and Privacy. Bus. Soc. (2017).
31. Evans, D.S.: The Online Advertising Industry: Economics, Evolution, and Privacy. J. Econ. Perspect. 23, 37–60 (2009).
32. Krishnamurthy, B., Wills, C.: Privacy diffusion on the web. In: Proceedings of the 18th international conference on World wide web. p. 541. ACM Press, New York, New York, USA (2009).
33. Falahrastegar, M., Haddadi, H., Uhlig, S., Mortier, R.: Anatomy of the Third-Party Web Tracking Ecosystem. arXiv. (2014).
34. Vallina-Rodriguez, N., Sundaresan, S., Razaghpanah, A., Nithyanand, R., Allman, M., Kreibich, C., Gill, P.: Tracking the Trackers: Towards Understanding the Mobile Advertising and Tracking Ecosystem. CoRR. abs/1609.0, 4–9 (2016).
35. Klein, S., Hüllmann, J.: Datenkapitalismus akademischer Wissenschaftsverlage. Wirtschaftsdienst. 98, 477–480 (2018).
36. Vissers, T., Nikiforakis, N., Bielova, N., Joosen, W.: Crying Wolf? On the Price Discrimination of Online Airline Tickets. In: 7th Workshop on Hot Topics in Privacy Enhancing Technologies. , Amsterdam, Netherlands (2014).
37. Schelter, S., Kunegis, J.: Tracking The Trackers: A Large-Scale Analysis of Embedded Web Trackers. In: Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016). pp. 679–682 (2016).
38. Roesner, F., Kohno, T., Wetherall, D.: Detecting and defending against third-party tracking on the web. Proc. 9th USENIX Conf. Networked Syst. Des. Implement. (2012).
39. Nikiforakis, N., Kapravelos, A., Joosen, W., Kruegel, C., Piessens, F., Vigna, G.: Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In: 2013 IEEE Symposium on Security and Privacy. pp. 541–555. IEEE (2013).

From Facets to a Universal Definition – An Analysis of IoT Usage in Retail

Clarissa Endress[1], Gregor Friedrich-Baasner[2], and David Heim[2]

[1] University of Wuerzburg, Chair of Business Information Systems, Wuerzburg, Germany
{clarissa.endress}@stud-mail.uni-wuerzburg.de
[2] University of Wuerzburg, Chair of Business Information Systems, Wuerzburg, Germany
{gregor.friedrich-baasner, david.heim}@uni-wuerzburg.de

Abstract. The term Internet of Things (IoT) is widely used and discussed. However, there is still no scientifically generally valid definition and therefore no uniform understanding of the term in practice. To counteract this, we present a possibility to limit the term with the help of applications in the IoT environment. Specifically, we analyze the use of IoT along the value chain of companies by conducting a structured literature analysis.

Keywords: Internet of things, applications, use cases, value chain, retail

1 Introduction

The Internet of Things describes the digitization and internet-capable integration of physical objects into the networked society [1]. In addition, the term includes things that not only collect local information autonomously, but also learn from it, adapt it and use it to interact with the world around them [2-4]. The aim of such devices is to create added value through appropriate services. In this case, the notion of IoT goes back to Kevin Ashton and the Auto-ID Center at the MIT. They designed a company-wide radio frequency identification (RFID) infrastructure in 1999 [5].

Today, there is an increasing number of previously non-digital products (e.g. light bulbs, bikes and watches) being equipped with sensing technology, application logic, and a network connection. Generally referred to as product service systems (PSS), they are defined as a combination of products and services. While these services initially only extended the functionality of a product [6], such systems may have inverse roles nowadays, in which the product is only a vehicle of the service. Their emergence has led to numerous areas of application in sectors like energy, healthcare, or home automation, changing the nature of business models in these industries [3]. IoT is also omnipresent within companies in all areas such as production and logistics. Through 2020, an estimated 26 billion smart objects will be installed, generating a market size of approximately $7.1 trillion [7-9].

Although IoT is omnipresent and a widely discussed multidisciplinary topic in various fields such as computer science and telecommunications, there is still no unified and valid definition of the term in the academic literature [10-12]. According to our

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

research, there is currently a different understanding of what is meant by IoT, which in turn makes it difficult to discuss. In order to counteract this situation, we have set ourselves the goal of developing a universally valid definition of IoT in science and practice. In particular, we want to counteract marketing terms such as Industry 4.0 and analyze the technical bases of the terms. From our point of view, IoT is nothing new, but merely the linking of existing technical components with corresponding business models. Some researchers have also noted a clear need for research in this area [1], [11], [12]. In order to define an object, it is necessary to become aware of all facets of it. For this reason, we have conducted an extensive investigation of existing applications of IoT. Since the variety of IoT applications ranges from music events to intelligent devices in the medical field, we have limited ourselves to the business process of value creation in a company in order to be able to make an abstraction on this basis [13], [14]. By identifying all facets of the IoT along the value chain, we try to investigate all similarities and then summarize them into a common definition. The paper on hand aims to present our research idea and the first results of the far-reaching literature analysis.

2 Theoretical Background

2.1 Research Methods

As part of this work and the associated search strategy, the literature analysis process was extended or modified in some places according to vom Brocke et al. [15] (cf. Figure 1). In addition to the aforementioned five-step search process, reference was also made to the information flow diagram of Frost and Lyons [16].

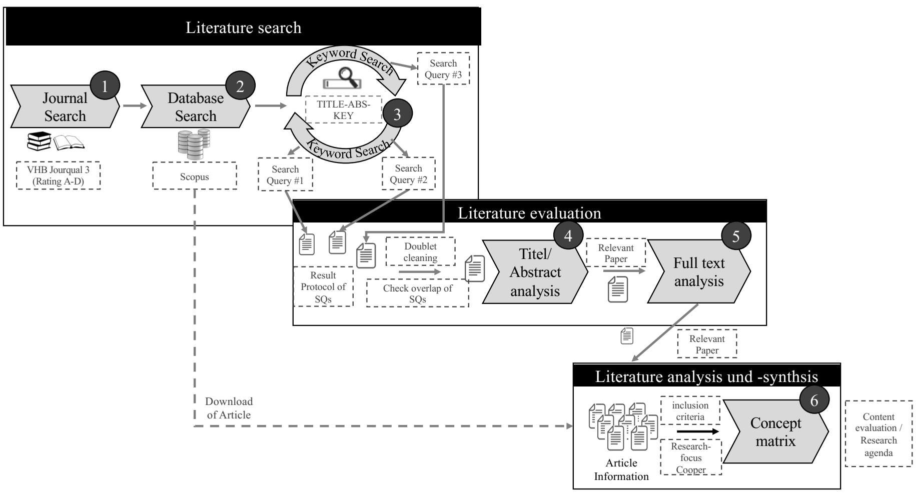


Figure 1. Approach to literature research

The figure illustrates the extended, six-stage search process which was carried out within the framework of this literature search. It should be emphasized that the keyword search consists of several iterations (phase 3). After defining the framework conditions within the first and second phase according to the research question, the actual search process can be started. The basis for this search is the so-called extraction log, in which all the necessary article information for each iteration is stored and which is used at the end for a final concept matrix. The matrix consists of articles, which are filtered out and the final search string can be worked out step by step. After the third phase has been successfully completed, the actual analysis steps (phase 4 & phase 5) can be started, which in this case do not differ from the procedure of vom Brocke et al. [15]. At the end, the result is the final extraction log, which in turn can be used as the basis for the concept matrix. It should also be mentioned that although forward and backward searches are relevant for the classic vom Brocke et al. [15] method, they were not part of the literature search strategy in the context of this work. The results of the individual phase and the number of identified contributions after the various search iterations are explained in more detail in the following chapter and evaluated accordingly.

2.2 Literature Review

As literature addressing definitions of IoT is very limited, we expand the focus of our research goal by looking at every facet of IoT. Due to the huge amount of literature concerning IoT and in line with scholars who recommend focusing on high quality outlets, we only considered articles published in outlets ranked (A+ to D) in the VHB-JOURQUAL3 [17]. For this purpose, the respective sub-ratings of the VHB were used as the basis for the search. We conducted a structured literature review following the methodology explained above. We used a search query consisting of three building blocks. On the one hand, these consisted of the IoT focus and the Value Chain focus with the respective key words following Porter [18], which can be seen in Figure 3, and on the other hand of the scientific outlets mentioned above. The keywords had to appear either in the title, abstract or keywords.

Our search resulted in an initial set of 596 papers. We analyzed the applicability of the cases presented in the papers by their title and abstract. The main reason for the exclusion of papers was that they either did not reveal any facets of IoT or they did not present any abstractable use cases for our research objective. The number of results was already reduced by more than half after this step. This analysis step can already be described as very accurate, because the subsequent full text analysis only classified 32 of the remaining articles as not relevant, leading to 225 articles for the content evaluation.

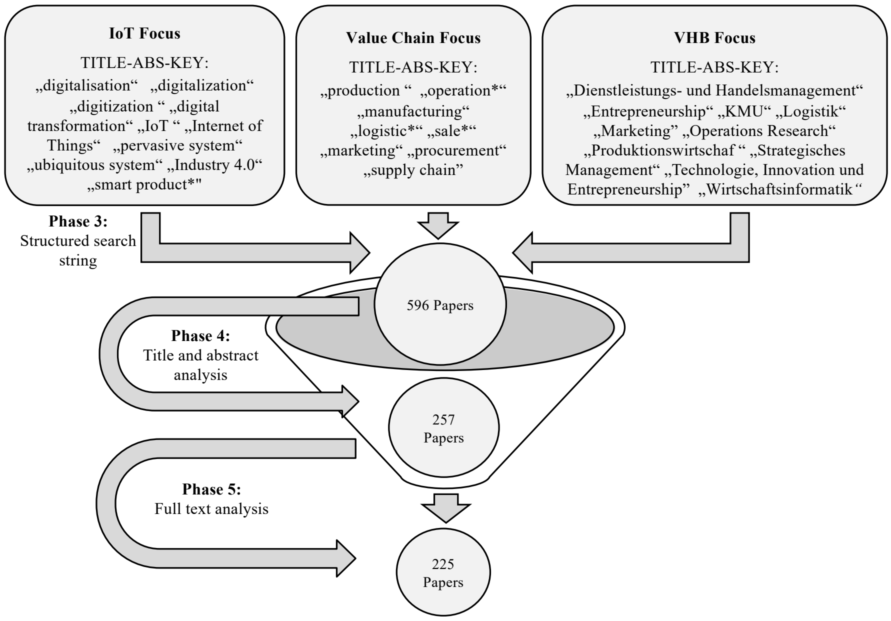
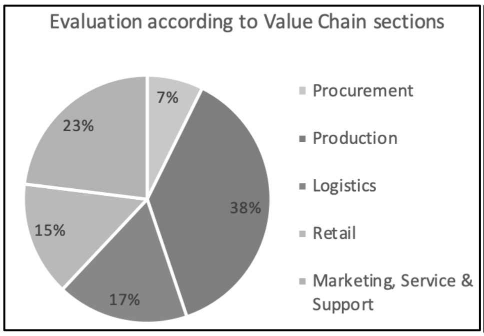


Figure 2. Literature analysis process

3 Evaluation

In this chapter, the content of the literature within the different concepts or business areas of the value chain is examined for concrete application scenarios of IoT. Therefore, the results are first clustered according to their respective value chain areas in order to then graphically analyze the individual technology approaches and their connections to each other. With regard to obtain a compact overview of the literature, we have not looked at each article individually, but have worked out the most frequently occurring application potentials of IoT. The following analysis is therefore based on exemplary IoT scenarios, which should not be regarded as exhaustive, but merely reflect a representative section of the literature. The focus lies on the technology drivers behind the recent transformation wave and their integration approaches among each other, since most technologies do not occur in isolation, but in interaction and therefore, compatibility is of crucial importance. However, before we go into the detailed evaluation, the following illustrations should give a first graphical overview of the article distribution per company division as well as the technology distribution. It might occur that an article deals with several or even all divisions.



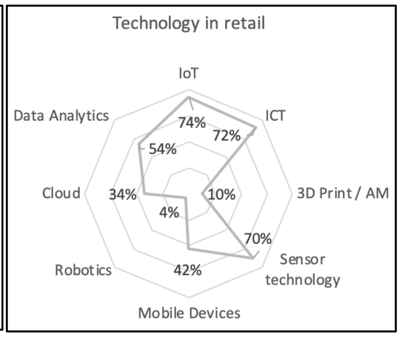


Figure 3. Evaluation by value chain area **Figure 4.** Technology distribution in retail

59 of the 225 identified articles deal with the IoT topic in the sales environment. In this context the concept of "smart retailing" was introduced, which refers to the use of retail IoT technologies to improve the quality of the shopping experience. In this scenario, smart and intelligent technologies are seen as enabling innovation and improving the quality of life of consumers [19]. How these technologies are reflected in the literature identified as relevant can be seen in figure 4.

It should be noted that in the case of sales-related processes, the smartphone is most frequently mentioned as IoT technology – compared to the processes in other value chain areas. Customers usually carry these devices with them and companies have the ability to access the display or payment function of such devices easily and integrate them into their business activities. The authors also see many application potentials for data analytics on the sales floor. For example, the product and customer data can be analyzed to dynamically adjust prices at any time. More details will be explained in the context of the appropriate subchapter. The evaluation of the literature has also shown that sensors and corresponding chips can be used in the retail context on different levels (product, shelves, cash machines, e.g.) or are already actively used. It is not surprising that robotics and additive manufacturing (AM) have a low number of hits because the overall evaluation across all value chain areas has already revealed this trend. In the case of retail processes, information and communications technology (ICT) can be seen as a basic module for all the other technologies because it provides the appropriate network or platform, integrates other IoT technologies and interfaces with other technologies or systems (n = 36).

On the one hand, smart products can provide support during the actual purchasing process (i.e. still on the sales floor), but on the other hand they can also be useful in the after-sales area and thus in the private domestic environment. In an article, the authors Miranda et al. [20] dealt, among other things, with sustainability considerations in the various phases of the life cycle of an intelligent product, and in this context can name some advantages in the area of product use and the end of product life. For example, after-sales maintenance and repair costs can be reduced through the use of e-services and remote diagnostics, thus increasing customer satisfaction [21]. If customers no longer need their products after a certain time, manufacturers can, for example, simply take back the old products, reconfigure them and then reuse or resell them [22]. Such functionalities can of course only be implemented to a limited extent with purely mechanical products, which is why the added value for smart products from the customer and company point of view can be seen in economic, environmental and social aspects [20].

Cloud Computing

Current trends in the Industry 4.0 research environment show cloud computing as one of the key technologies for IoT applications [23]. Especially in the context of sales processes, the provision of digital services requires the combination of cloud content, a physical product with embedded hardware and software, and a link channel to provide the services to the customer [24]. Such a service provision for the customer can be implemented, for example, by a cloud-based smartphone-app like the start-up BEXT360 within the coffee industry [25] or by integrating smart refrigerators into a cloud-based ecosystem [22].

ICT

Companies such as Google, Apple and Microsoft have been competing in the area of Enterprise System Software (ESS) for many years. The shift from ESS to web services, cloud computing and PSS, rather than installed components, provides future researchers new possibilities [26]. In addition to platform strategies, cyber-physical systems (CPS) also play an important role. A mechatronic system that uses CPS is an intelligent product due to its composition that integrates embedded systems and enables communication with other smart products. In general, a product can be smart if it uses ICT to improve its functionality and productivity. This is how the concept of PSS emerged, which integrates products and e-services into individual solutions for customers [20]. In the context of radio-frequency identification (RFID) different information systems or platforms may be used, depending on the industry or technology focus. One article speaks of a so-called EPICS-System (short for electronic product code information service). In this case, a GS1 standard automatically detects the movement of RFID tags and processes events according to predefined policies [27]. The result can be regarded as an RFID-based self-contained information system, which brings enormous advantages on the customer and company side.

Data Analytics

The authors Simchi-Evi and Wu [28] have analyzed the potentials of Big Data analysis in the context of price optimization models and price strategy adjustments, because not until now the retail sector has been in such a complex and competitive environment as in the current digitalization age. For example, based on product images and supplier information, customers in the Netherlands have the opportunity to bid online using a virtual auction watch [29]. Other authors have also focused on pricing strategies and models in the context of Big Data analysis tools. For example, in his article Weber [30] dealt with artificial intelligence systems in connection with sharing markets. By embedding intelligence in products, businesses and individuals, a so-called collaborative consumption arises. These recent developments have also encouraged some players in the automotive industry to develop new sharing solutions for urban mobility. The business initiative ENJOY, which was developed through a joint venture between Fiat Chrysler and ENI, can be shown as an example. With their smartphones, customers can quickly find, book and activate the desired vehicle and pay for the service based on their actual usage [31]. Amazon has tremendous potential to customize the value proposition for its customers. Through intelligent algorithms, Amazon knows where its customers live, what they have bought and what they might be interested in [32]. Nevertheless, companies such as Google or Ebay have also discovered these analytical capabilities and offer intelligent marketplace platforms as a service [33].

Sensor technology

The first study, which was considered relevant in the field of sensor technology, examines the needs of older consumers in the food sector and their implications for the food supply chain in a more holistic demand chain approach [34]. As a central result of this study, the following could be stated: Although the aging society also poses new challenges for the sales process, it can be positively counteracted by the use of IoT technologies. Among other things, the authors mention RFID technology, because RFID can significantly simplify the purchasing process on the customer side. Another author differentiates between two distinct RFID-based product applications on the sales area: (1) RFID as a central feature of the object itself and (2) RFID as a marking for the respective objects. The first group represents a central benefit for the customer, whereas the second option brings significant benefits to producers or companies [35]. Practical examples in the context of the first deployment scenario include a smart mirror, such as one from Panasonic, which is currently on the market, and OAK's smart changing room. Smart shelves, on the other hand, can be assigned to both RFID product applications. For example, Shelf Vending refrigerators are equipped with a system that recognizes and monitors the quantity and type of food being stored, so it automatically detects when consumers or employees remove products from the shelves [22]. No matter whether RFID chips are used on products, shelves or mirrors, the advantages for the authors are always the automation, the clear identification as well as the additional sources of information and the future will enable an easy customer and product profiling.

Mobile computing

Front-end digitization enables completely new types of customer interaction, such as creating self-service touch points with personal digital assistants, tablets or smartphones [36]. The first relevant articles deal with smartphones in the context of mobile coupons [19], [37]. The remaining articles all deal with the same application: smartphones as product information display. Quick response (QR) codes on the products give customers access to product information or product reviews directly on their smartphone display [22], [34]. Most common are applications that are not limited to providing product-related information, but also allow the sharing of product images and experience within a public network [19], [23]. Companies like Tesco and Amazon have taken advantage of the possibility to scan products via QR-codes and install appropriate applications within their business models [22]. Mobile payment systems have also become a hotspot for digital innovations. Smartphone manufacturers, telecommunications operators, payment service provider, software companies and start-ups are increasingly entering the payment market. We are not only talking about well-known companies like Google, Facebook or Apple, but also early-payment entrepreneurs such as Square, Paypal and Izettle [38]. These developments should significantly improve and automate the entire purchasing process – from information provision to payment process – on the customer and company side. Consequently, smartphones can be part of the IoT in various ways and making life easier for customers, companies and manufacturer.

AM

One of the articles considered the methodology of rapid prototyping (RP) as relevant in the context of AM. RP is currently mainly presented in the areas of production or research and development. According to the authors, however, the technology is being used more and more within the sales area. For example, RP can be used for product presentation to demonstrate the products in real. The big advantage is that the technology provides a prototype without long production time and allows to be modified several times, depending on the customers' needs. This deployment scenario is only possible due to the latest developments in machine learning and artificial intelligence techniques [39]. This shift of the product design phase to the customers can be seen as a decentralization approach, which creates a so-called temporarily supply chain. As a result, organizations are able to adopt ever-changing businesses and customer needs, without significant costs and time penalties [23].

Robotics

Robotics, on the other hand, currently has little acceptance in the context of sales processes, as only one article deals with this kind of IoT technology. The authors focus on a special kind of robotics, namely the so-called wearable robotics. Wearable robotics is a new field of research, focused primarily on the development of intelligent assistive devices and exoskeletons, which support human movement or enhance human performance. In this case, the robots consist of a series of actuators and sensors and require the development of advanced software and algorithms to process the data and control the system. Currently smart auxiliary devices (e.g. for arms) are commercially available and used by logistics providers. They can support different kinds of physical tasks, e.g. lifting and picking heavy food packs. Nevertheless, research on exoskeletons is still at an early stage and primarily focuses on military or medical applications [22].

In the context of sales-related business processes, all technology directions could be identified as relevant and explained using a wide variety of practical examples. One the one hand many authors have dealt with same or similar technology approaches and, on the other hand, some niche potentials have been addressed in some articles. As a result, there are smart products that can be found in multiple industries or business sectors – such as the location function of smartphones and corresponding apps – and those that are more specially designed for their use.

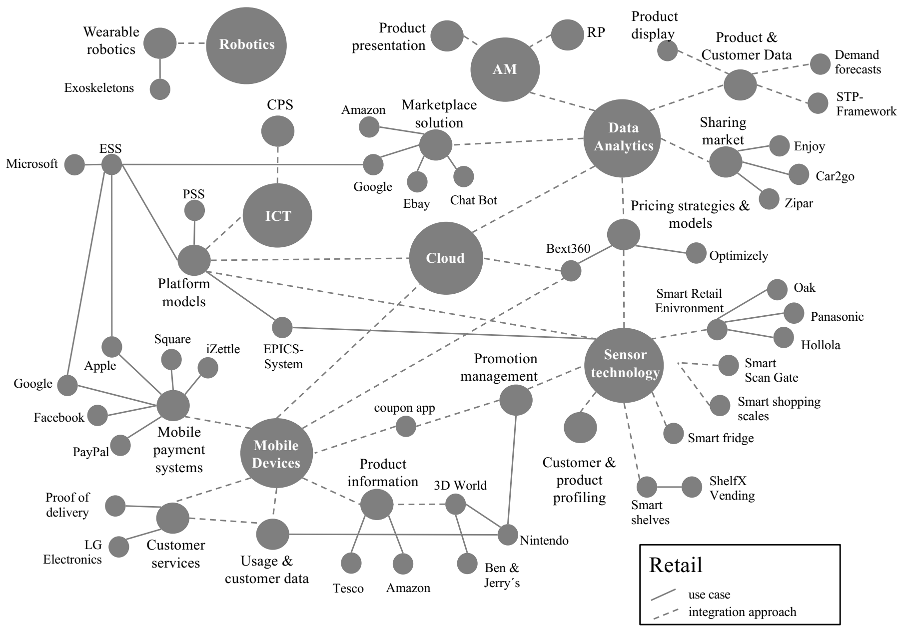


Figure 5. IoT application scenarios in retail

4 Conclusion and Future Work

The paper provides practical insights from manufacturers in various industries who have already had advanced experience with the implementation of IoT technologies. These insights can be seen as a source of learning effects for other companies or researchers, which plan to implement IoT solutions in their value chain.

From the scientific literature a clear mandate for further research of IoT can be derived [11], [12]. This also serves as motivation for the research approach presented below.

The results of the literature analysis presented here serve as an impetus for the research project. Use cases were extracted from the multitude of literature. Furthermore, on the basis of these, properties and facets of IoT can be subsumed. The far-reaching literature base then also serves further to create a taxonomy, if one follows Nickerson et al. [40].

The methodological approach by Nickerson et al. [40] can be used to structure IoT services based on conceptual and empirical observations. A taxonomy is defined as a set of dimensions, each consisting of different, mutually exclusive, and collectively exhaustive characteristics such that each object under consideration has one characteristic for each dimension [40]. Taxonomies play an important role in structuring novel terms and technologies to provide a deeper understanding of emerging fields of research and application [41]. Haas et al. [42] for example developed an empirical taxonomy in the field of crowdfunding intermediaries. Through the development of this taxonomy and the subsequent analysis of the collected data, they identified three archetypes of crowdfunding intermediaries.

Through collecting empirical and conceptual data in the area of IoT, common characteristics and dimensions can be identified. Subsequently, a common definition can be derived from the general characteristics.

Despite the fact that the literature analysis was conducted with a methodical rigor to improve the representativeness, the results should not be generalized to other empirical areas. There are still the natural limitations: (1) restriction on German and English language, (2) selected keywords search, (3) the selected database and the number of articles that it contains, and (4) the geographic and content characteristics of the articles identified as relevant, making it difficult to generalize to all sectors and countries.

As future work we will employ a seven-step method as proposed by Nickerson et al. [40]. For this purpose, we will collect a unique empirical data set of IoT applications and define meta-characteristics. These characteristics serve as a reference point for the dimensions and characteristics which are to be identified. Furthermore, we use the suggested subjective and objective ending conditions to determine when the taxonomy is robust. At this stage, we are able to derive a common definition of the term IoT.

Through the above-mentioned process, we are seeking for successful results within our future research.

References

1. Rosemann M.: The Internet of Things: new digital capital in the hands of customers. In: Business Transformation Journal, Vol. 9, pp. 6-15 (2014)
2. Ashton, K.: That "Internet of Things" Thing. RFID Journal (2011)
3. Atzori, L., Iera, A., & Morabito, G.. The Internet of Things: A survey. Computer Networks, 54(15), 2787–2805 (2010)
4. Weiser, M.: The computer for the 21 st century. ACM SIGMOBILE Mobile Computing and Communications Review, 3(3), pp. 3–11 (1999)
5. Hsu C.-W.; Yeh C.-C.: Understanding the factors affecting the adoption of the Internet of Things. In: Technology Analysis and Strategic Management, Vol. 29, No. 9, 1089-1102 (2017)
6. Baines, T. S., Lightfoot, H. W., Evans, S., Neely, A., Greenough, R., Peppard, J., … Wilson, H.: State-of-the-art in product-service systems. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 221(10), pp. 1543–1552 (2007)
7. Fleisch, E.: What is the Internet of Things? An Economic Perspective. Economics, Management & Financial Markets, 5(2), pp. 125–157 (2010)
8. Macaulay, J., & Kückelhaus, M.: Internet of Things in Logistics - A collaborative report by DHL and Cisco on implications and use cases for the logistics industry. DHL Trend Research / CISCO Consulting Services (2015)
9. Wortmann, F., & Flüchter, K.: Internet of Things: Technology and Value Added. Business & Information Systems Engineering, 57(3), pp. 221–224 (2015)
10. Bandyopadhyay, D., & Sen, J.: Internet of Things: Applications and Challenges in Technology and Standardization. Wireless Personal Communications, 58(1), pp. 49–69 (2011)

11. Boos, D.; Guenter, H.; Grote, G.; Kinder, K.: Controllable accountabilites: the Internet of Things and its challenges for organisations. In: Behaviour & Information Technology, 32(5), pp. 449-467 (2012)
12. Kees, A., Oberländer, A. M., Röglinger, M., & Rosemann, M.: Understanding the Internet of Things: A Conceptualisation of Business-to-Thing (B2T) Interactions. In 23rd European Conference on Information Systems (ECIS) (2015)
13. Friedrich-Baasner, G.; Heim, D.; Hauser, M.; Gramer, W. & Winkelmann, A.: Improving the processes of music events with radio frequency identification. In: IEEE International Conference on RFID Technology & Application (RFID-TA), Warsaw, pp. 119-124 (2017)
14. Angst, C. M., and Ritu, A.: Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. In: MIS Quarterly. Vol. 33, No. 2, pp. 339-70 (2009)
15. Vom Brocke, J.; Simons, A.; Niehaves, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstruction the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of the 17. European Conference on Information Systems (ECIS), Verona, pp. 1-12 (2009)
16. Frost R.; Lyons K.: Service Systems Analysis Methods and Components: A Systematic Literature Review. In: Service Science, Vol. 9, No. 3, pp. 2019-234 (2017)
17. Rowley J.; Slack F.: Conducting a literature review. In: Management research news (27:6), pp. 31–39 (2004)
18. Porter M.: Competitive advantage – Creating and Sustaining Superior Performance. In: The Free Press, New York, pp. 36-43 (1985)
19. Willems K.; Smolders A.; Brengman M.; Luyten K.; Schöning J.: The path-to-purchase is paved with digital opportunities: An inventory of shopper-oriented retail technologies. In: Technological Forecasting and Social Change, Vol. 124, pp. 228-242 (2017)
20. Miranda J.; Pérez-Rodríguez R.; Borja V.; Wright P.K.; Molina A.: Sensing, smart and sustainable product development (S3 product) reference framework. In: International Journal of Production Research, pp. 1-22 (2017)
21. Lee J.: Smart products and service systems for e-business transformation. In: International Journal of Technology Management, pp. 33-38 (2003)
22. Vanderroost M.; Ragaert P.; Verwaeren J.; De Meulenaer B.; De Baets B.; Devlieghere F.: The digitization of a food package's life cycle: Existing and emerging computer systems in the logistics and post-logistics phase. In: Computers in Industry 87, pp. 15-30 (2017)
23. Barata J.; Rupino Da Cunha P.; Stal J.: Mobile supply chain management in the Industry 4.0 era: An annotated bibliography and guide for future research. In: Journal of Enterprise Information Management, Vo. 31, No.1, pp. 173-193 (2018)
24. Vendrell-Herrero F.; Bustinza O.F.; Parry G.; Georgantzis N.: Servitization, digitization and supply chain interdependency. In: Industrial Marketing Management, pp. 1-13 (2017)
25. Kshetri N.: 1 Blockchain's roles in meeting key supply chain management objectives. In: International Journal of Information Management 39, pp. 80-86 (2018)
26. Chellappa R.K.; Sambamurthy V.; Saraf N.: Competing in crowded markets: Multimarket contact and the nature of competition in the enterprise systems software industry. In: Information Systems Research 21, pp. 614-630 (2010)
27. Han W.; Gu Y.; Wang W.; Zhang Y.; Yin Y.; Wang J.; Zheng L.-R.: The design of an electronic pedigree system for food safety. In: Information Systems Frontiers 17, pp. 275-287 (2015)
28. Simchi-Levi D.; Wu M. X.: Powering retailers' digitization through analytics and automation. In: International Journal of Production Research, Vol. 56, pp. 809-816 (2017)

29. Verdouw C.N.; Beulens A.J.M.; Reijers H.A.; Van Der Vorst J.G.A.J.: A control model for object virtualization in supply chain management. In: Computers in Industry, pp. 1-16 (2015)
30. Weber T. A.: Smart Products for Sharing. In: Journal of Management Information Systems, Vol. 34, No. 2, pp. 341-368 (2017)
31. Ardolino M., Rapaccini M., Saccani N., Gaiardelli P., Crespi G., Ruggeri C.: The role of digital technologies for the service transformation of industrial companies. In: International Journal of Production Research, Vol. 56, No. 6, pp. 2116-2132 (2018)
32. Balmer und Yen D.A.: The Internet of total corporate communications, quaternary corporate communications and the corporate marketing Internet revolution. In: Journal of Marketing Management, Vol. 33, No. 1/2, pp. 131-144 (2017)
33. Varadarajan R.; Yadav M.S.: Marketing Strategy in an Internet-Enabled Environment: A Retrospective on the First Ten Years of JIM and a Prospective on the Next Ten Years. In: Journal of Interactive Marketing, Vol. 23, pp.11-22 (2009)
34. Oeser G.; Aygün T.; Balan C.-L.; Corsten T.; Dechêne C.; Ibald R.; Paffrath R.; Schuckel M.T.: Implications of the ageing population for the food demand chain in Germany. In: International Journal of Retail and Distribution Management, Vol. 46, No. 2, pp. 163-193 (2018)
35. Slettemeås D.: RFID - The next step in consumer-product relations or orwellian nightmare? Challenges for research and policy. In: Journal of Consumer Policy 32, pp. 219-244 (2009)
36. Coreynen W.; Matthyssens P.; Van Bockhaven W.: Boosting servitization through digitization: Pathways and dynamic resource configurations for manufacturers. In: Industrial Marketing Management, Vol. 60, pp. 42-53 (2017)
37. Beeck I.; Toporowski W.: When location and content matter: effects of mobile messages on intention to redeem. In: International Journal of Retail and Distribution Management, Vol. 45, No. 7/8, pp. 826-843 (2017)
38. Graupner E.; Maedche A.: Process digitisation in retail banking: An empirical examination of process virtualization theory. In: International Journal of Electronic Business, Vol. 12, No.4 , pp. 364-379 (2015)
39. Syam N.; Sharma A.: Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. In: Industrial Marketing Management, Vol. 39, pp. 80-92 (2018)
40. Nickerson, R. C., Varshney, U., & Muntermann, J.: A method for taxonomy development and its application in information systems. European Journal of Information Systems, 22(3), pp. 336–359 (2013)
41. Bühler, R.: Standards in Disruptive Innovation: Assessment Method and Application to Cloud Computing (2015)
42. Haas, P., Blohm, I., & Leimeister, J. M.: An Empirical Taxonomy of Crowdfunding Intermediaries. In: Proceedings of the 35. International Conference on Information Systems (ICIS), Auckland, pp. 1-18 (2014)

Is the Technostress Creators Inventory Still an Up-To-Date Measurement Instrument? Results of a Large-Scale Interview Study

Thomas Fischer[1], Alexander Pehböck[1], and René Riedl[1,2]

[1] University of Applied Sciences Upper Austria, Steyr, Austria
{thomas.fischer, rene.riedl}@fh-steyr.at
Alexander.pehboeck@students.fh-steyr.at
[2] University of Linz, Austria

Abstract. Self-report measures are of great importance for technostress research and particularly the Technostress Creators Inventory has been amongst the most frequently applied measurement instruments since its inception in 2008. As the technological environment has progressed since then, we investigated whether this inventory still completely represents the phenomenon of technostress today. We conducted interviews with 75 individuals in four companies on their technostress experiences. We asked them about their personal experiences with eight stressor categories in the workplace, the original five dimensions of the Technostress Creators Inventory (techno-overload, techno-invasion, techno-complexity, techno-insecurity, and techno-uncertainty) and additional three categories (techno-unreliability, IT-based monitoring, and cyberbullying). We found that techno-insecurity was the least prevalent stressor category throughout all companies, while techno-unreliability, one of our new categories, was the most prevalent stressor. Based on this evidence, we argue that a revised inventory is urgently needed to guarantee content validity of technostress measurement in future studies.

Keywords: Technostress, Stressors, Measurement, Self-Report, Interviews

1 Introduction

Technostress is a phenomenon that arises from the individual use of information and communication technologies (ICT) [1]. The body of knowledge concerning this dark aspect of IS use has grown rapidly in recent years (e.g., [2–4]) and its investigation relies strongly on the use of self-report instruments [5]. The main self-report instrument used for technostress measurement is the Technostress Creators Inventory (TCI) introduced by Ragu-Nathan et al. [1]. This instrument conceptualizes technostress as a superordinate construct with five first order constructs each having three to five reflective indicators. Specifically, the instrument conceptualizes technostress as being manifested in the five dimensions techno-overload (*too much*),

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

techno-invasion (*always connected*), techno-complexity (*difficult*), techno-insecurity (*uncomfortable*), and techno-uncertainty (*too often and unfamiliar*) [6].

As this instrument celebrated its tenth anniversary recently, we wanted to know whether it still offers the necessary content validity to capture technostress in a rapidly evolving technological environment. More specifically, we focus on the instrument's first-order constructs, and in line with MacKenzie et al. [7] (p. 304) pose the question: *Are the [first-order constructs] as a set collectively representative of the entire content domain of the construct*? Answering this question is important as it may reveal the need for a revision of the instrument. Such a revision could further improve the diagnostic abilities of the Technostress Inventory, which are crucial to investigate the phenomenon and its detrimental effects efficiently (e.g., to avoid missing important facets that may have emerged during the last ten years).

An initial indication for the need to revise the instrument can be found in previous technostress research. In a number of studies, only a subset of the original five technostress creators were used. For example, Tarafdar et al. [8] used only four of five creators (i.e., overload, invasion, complexity, insecurity) or D'Arcy et al. [9] used only three of five (i.e., overload, complexity, uncertainty). Importantly, it is also critical to assess whether further dimensions are necessary to completely represent the technostress phenomenon. Hence, in addition to the five existing technostress creators we added three additional dimensions (i.e., unreliability, monitoring, cyberbullying). Based on a series of interview studies in four companies in Austria, we investigated whether the original five Technostress Creators are sufficiently representative of the entire content domain. Regularly checking this property is critical as the technological environment that causes technostress evolves constantly. While technostress was originally caused by automation in the workplace, it evolved into a computer-related problem later [10], and may again manifest differently nowadays.

In Section 2, we introduce the TCI in more detail and why we expect that a measurement instrument so tightly connected to the technological domain would benefit from an update. In Section 3, we present an overview of our research methods including the interview guide that was used for all 75 interviews, a characterization of the companies that were used as the sample for our investigation, and a description of the data analysis process (including the coding process and interrater agreement computation). In Section 4, we present the results of our investigation. Finally, in Section 5, we close with conclusions, including the implications of our findings.

2 Theoretical Background

The Technostress Creators are not the first inventory that was developed to measure stressors related to ICT. Another noteworthy, but far earlier inventory was developed by Richard Hudiburg, called the Computer Hassles Scale [11], [12]. This self-report inventory serves as a good example for the necessity to regularly update measurement scales related to technology as Hudiburg himself emphasized the need to make an update of his original instrument within only a few years. This was necessary as the original instrument included items that were related to specific technologies that were somewhat new and stress-inducing at the time (e.g., automatic bank tellers or library computer scanners) [12], but became common technologies fast and were therefore removed as potential sources of stress in a later version of the inventory [11].

Developments of this kind have also been acknowledged by Ragu-Nathan et al. [1] who stated that "…modern ICTs have changed the work environment and culture." (p. 418). We endorse this statement and see no reason why this development would have stopped in the last decade. Additional evidence for the need for such an update was provided by Fischer and Riedl [5], who argued that the current conceptualization of technostress in the inventory may not fully encompass common types of technologies found in the workplace (e.g., as evidenced by the need to adapt techno-overload for the context of social media).

For our investigation, in addition to the original five technostress domains, we selected three potential technology-related stressors that have received significant interest by both researchers and practitioners, namely unreliability of technology (e.g., unexpectedly long system response times), monitoring via technology (e.g., electronic performance monitoring), and cyberbullying (e.g., impolite remarks made through electronic communication media). It has to be noted, that although important, these categories reflect a convenience sample of technology-related stressors which could be added to the Technostress Creators Inventory. In addition to this explorative investigation, therefore, a more systematic quantitative investigation is needed to further assess whether these categories, or other eventually also other categories, can strengthen the factorial structure of the inventory.

Unreliability. Reliability is an important feature of information systems that may never be completely achievable as stated by Butler and Gray [13] who expected that "…many important systems are not (and perhaps cannot be) inherently reliable." (p. 212). It has also been shown that system reliability is an important stress-related feature of ICT. For example, Ayyagari et al. [2] showed that system reliability is negatively related to perceived work overload, and Riedl et al. [14] demonstrated the detrimental effect of the malfunction of a webshop on individual stress physiology (i.e., a significant increase in the levels of the stress hormone cortisol).

Monitoring. Employee dissatisfaction related to electronic performance monitoring has been an early topic in technostress research. For example, in a nation-wide survey study amongst 762 employees in the USA, Smith et al. [15] found that monitored individuals report higher levels of job-related stress and physical strains independent of profession (e.g., sales representatives, but also clerks). In more recent years, workplace monitoring has also received considerable public interest, particularly as the threat for individual data privacy has become a more important focus [16], [17].

Cyberbullying. A more recent phenomenon in the workplace is related to the use of electronic communication media for negative behaviors (e.g., offensive comments and insults). Thus far, cyberbullying has mostly been focused on in the context of school children and young adults[1]. Using a sample of white-collar workers, Snyman

[1] https://www.nytimes.com/topic/subject/cyberbullying [07/17/2018].

and Loh [18] more recently confirmed the positive relationship between cyberbullying experiences at work and perceived daily stress in a survey study.

3 Methods

We investigated the prevalence of technological stressors through 75 interviews in four companies. Specifically, we constructed an interview guide, collected data, extracted relevant data in the transcripts (i.e., experiences related to technology-induced stress), coded the interview material, and summarized the data [19], [20]. We chose qualitative interviews because they allow us to grasp a wide area of individual experiences and give us the opportunity to make clarifications wherever needed (e.g., related to the definition of "stress", which can have different meaning across individuals).

Interview Guide. We first constructed an overall interview guide (originally in German) that was then slightly adapted for each company context (i.e., using a pre-test with 2 to 3 employees and adaptations in wording to increase comprehension by our interview partners). Keywords and descriptions for the original Technostress Creators were taken from Tarafdar et al. [6] and the remaining category descriptions were constructed in a comparable fashion:

- Unreliability (*Too unstable*): IS users are confronted with system malfunctions and unexpected system behaviors.
- Monitoring (*Controlled*): IS users feel constant levels of pressure as their behaviors can be tracked by technology and hence evaluated by other individuals.
- Cyberbullying (*Vulnerable*): IS users feel threatened by social misconduct facilitated by communication technologies and social media.

Sample and Data Collection. The sample included two small family-owned businesses and two medium to large sized organizations in Austria with up to several thousands of employees. In each organization, we aimed for a sample of individuals that was representative of the departments of the local branch of the organization or the organization as a whole (e.g., focusing on the largest departments). Participation in the 75 face-to-face interviews was voluntary. All interviews were recorded based on written consent given by the interview partners. An overview of the sample characteristics and the collection periods can be found in Table 1.

Each interview started with the interviewer first explaining the general topic of technostress, then asking the participant to sign a written consent form, and afterwards the technostress categories were briefly explained before the interviewee was asked to tell the interviewer about personal experiences with each technostress category they had at work. For the technostress categories, we used the same order every time (i.e., overload, invasion, complexity, insecurity, uncertainty, unreliability, cyberbullying, monitoring, other stressors); it has to be noted though that at each point interviewees were able to talk about their experiences related to each category. Interview lengths ranged from 9 to 78 minutes with a slight variance between organizations (company A: 18 to 70 minutes; company B: 14 to 78 minutes; company C: 9 to 26 minutes; company D: 13 to 53 minutes). The particularly short length of interviews in company C could be attributed to efficient coping with technostress, facilitated by a high level of decision latitude in the organization (e.g., individuals can switch easily between activities and are not mandated to be available for further tasks as soon as they leave the workplace).

Table 1. Sample characteristics and overview of collection periods

ID	*Type of business*	*N*	*Age Median*	*Collection Period*
A	Software development	15 (2 f)	42	03/16/2017-05/04/2017
B	Passenger transportation	21 (8 f)	50	12/06/2017-03/12/2018
C	Publishing	24 (20 f)	38.5	12/14/2017-03/21/2018
D	Petrochemical production	15 (2 f)	43	11/20/2017-03/28/2018

Content Analysis and Interrater Agreement. The interviews were transcribed by the interviewers and potential experiences of technostress (e.g., instances were individuals showed annoyance or anger related to technology) were highlighted in each transcript using MAXQDA (version 11).

A total of 380 text passages related to technostress were then coded by the first and second author independently, using the stressor categories. Cohen's Kappa was used as a measure of interrater agreement [21]. Overall, a Cohen's Kappa of .84 was achieved, which indicates a strong agreement between the coders [22], with a range of .83 (company C) to .87 (company B). The remaining text passages, for which no initial agreement was achieved, were then discussed by the first and second author until a consensus was reached.

4 Results

In the following sections, we present insights for the original TCI and our additional stress categories, including concrete quotations from the interviews (literally translated from German). In addition, for each stressor we report on the share of individuals (rather than the absolute number, because of differences in samples sizes) who have had experiences with it before as well as an average (AVG) across all organizations (please note that this average is not weighted based on the sample size per company).

4.1 Technostress Creators Inventory (TCI)

From our investigation, we can conclude that there is a large variance in the share of individuals who had stressful experiences with techno-overload and techno-invasion across companies, while that share is widely consistent for techno-complexity, techno-insecurity, and techno-uncertainty.

Techno-Overload (A: 93%; B: 52%; C: 42%; D: 80%; AVG: 67%). In this category, we found a large discrepancy between companies, with the software developer (A) having the highest share of individuals who had experienced techno-overload before, while individuals in the transportation company (B) and the publishing company (C) had far less experience with this kind of stressor. In company A, most of the interviewed employees have short-term problem solving as part of their job responsibilities, which could explain the high prevalence of overload in this context. Particularly the dependence on communication technologies in some cases leads to a strong need to keep a focus on inboxes, with one software developer (A01, male, 44 years old) stating that: "*My main medium is e-mail. Within that medium, it is absolutely massive. Practically speaking, I spend about 80% of my time with e-mails*."

In the cases of companies B and C, this problem is far less prevalent and decision latitude seems to have an impact on this matter and in particular timing control [23], which is evident in the statement of one member of the management in company B (B01, male, 52 years old): *"On one hand, I think that [ICT] is stress-reducing, because it lets me know what is still to do. Back in the day, I had to wait until the evening until I received a long list of tasks for the next day. Now I know exactly what to do and can schedule ..."*

Techno-Invasion (A: 73%; B: 38%; C: 4%; D: 40%; AVG: 39%). In this category, we find more support for the assumption that timing control may have a strong influence on technostress perceptions. One of the manifestations of this relationship can be found in company A, were individuals have to cope with rigorous customer demands and a large variety of customized systems. A member of the support team (A02, male, 42 years old) addresses this point directly, stating that: "*As soon as I am awake, I basically look into it all the time and react to them [i.e., e-mails]. Because there is an implicit demand and our customers get really angry when they write an e-mail on Sunday and do not get a response."*

This pressure, which takes timing control away from the individual, as quick responses and solutions are needed, is far less of a problem in company C, as indicated by an employee who is responsible for managing online media outlets of the company (C01, female, 30 years old): "*Generally, if somebody is calling me in the office, I am taking the call. If this person then starts to chat away, for which I might not have time at work, I try to reschedule the call towards the evening*."

Techno-Complexity (A: 40%; B: 38%; C: 46%; D: 53%; AVG: 44%). Being overwhelmed by the possibilities and functionalities of ICT was a recurring problem in all companies, mostly caused by large numbers of different systems or a wide variety of different use cases that were not all needed regularly. One member of the research and development department in company D (D01, female, 39 years old) pointed out that there is often a lack of time to get used to operate complex systems, stating that: "*For example SAP, something like that is extremely complex. ... if you find the time to use it and get used to it… then you get able to handle such a system. Yet, that time is often missing.*"

Techno-Insecurity (A: 13%; B: 33%; C: 8%; D: 0%; AVG: 14%). The level of perceived threat to one's job security and potential changes to the job profile was generally low across all organizations. The highest prevalence of this stressor was in company B, with bus drivers fearing that they might get replaced someday. As one bus driver (B02, male, 55 years old) puts it: "*When the bus is controlled by computers someday, then I will not sit in it anymore. This will definitely become a reality, but I* *will not experience that during my time. We do not have to discuss that, this will definitely come to pass!*"

Techno-Uncertainty (A: 53%; B: 52%; C: 42%; D: 53%; AVG: 50%). As a stressor, the prospect of unexpected and rapid changes in the used technologies was comparable in its frequency to techno-complexity. There were also no main differences between companies in this domain, with most individuals reporting stress due to the uncertainty whether they would be able to use systems again after some sort of update. What is interesting though, is that the prospect of changes themselves is not that stressful, but what may come after, as expressed by a customer support employee in company A (A03, male, 27 years old): "*In our case, there are constant changes and new additions to programs. In many instances I do not really know if some added part works already or not*."

4.2 Techno-Unreliability

Unreliability of systems turned out as the most frequently cited stressor (A: 80%; B: 76%; C: 75%; D: 67%; AVG: 74%). In some cases, smaller hassles were reported as stress-inducing as they occurred frequently and therefore led to a chronic experience of stress. For example, a software developer in company A (A04, male, 37 years old) reported on his constant problems with a slow Internet connection: "*This is amongst our biggest topics – the stability of the Internet connection. ... There is nothing worse than hitting a button and there is a delay until it loads. Or when you try to scroll down on your monitor and you can watch it go down very slowly*."

In other cases, singular, non-frequent instances of unreliability, sometimes with serious consequences, were reported, such as lost work as reported by a member of the sales department in company C (C02, female, 54 years old): "*When an order gets lost or, in the case of e-mail, when I have written a comprehensive offer, was almost done and then it suddenly gets lost. That does not make you happy…*"

In each case, ICT behaved differently from what individuals expected and therefore caused anger or insecurity. In addition, comparable to techno-uncertainty, there is a spiral of problems that is related to unreliability. While slow response times themselves may not be problematic, these delays may lead to work overload as the current task cannot be handled sufficiently fast. Hence, time pressure is an important contextual factor for unreliability-related stress, just as the job-related dependence on technology [24]. This was indicated by a bus driver in company B (B03, male, 50 years old) who stated: "*Simply, because you are dependent on that stuff and in a crucial moment it often does not work. That is pure stress! When your navigation device does not work and you are not sure where to go*."

4.3 Monitoring

Comparable to techno-overload, there was also a large variance of the prevalence of experiences with monitoring between companies (A: 53%; B: 43%; C: 0%; D: 27%; AVG: 31%). While none of the individuals in company C had any stressful experiences with monitoring, more than half of the individuals in company A had made such experiences before. A potential connection here can be drawn between these observations and the level of control that is exerted in each company, for varying reasons. For example, regulatory demands (e.g., documentation of working hours) were often quoted as potential reasons for monitoring, which although often acknowledged as reasonable, nonetheless led to substantial stress. In company B, one of the bus drivers (B04, male, 58 years old) argued: "*I do not find it distracting, but not right, as we are completely supervised. It does not matter who it is ... they know at all times exactly where I am.*"

In the same vein, a customer support employee in company A (A05, female, 23 years old) stated: "*What I fear is that everything can be checked. What I do, what I say ... internal support can check everything. ... That leads to ... pressure*."

Monitoring and data collection are also often accompanied by a lack of transparency on what types of data are collected and why they are collected. Even if being monitored was not directly stress-inducing, the lack of transparency itself led to concern in individuals, which was particularly expressed by individuals in company A and D, who are knowledge workers handling extensive amounts of data. For example, a software developer in company A (A01, male, 50 years old) expressed particular concern stating: "*You do not feel threatened ... it is more of a concern due to the uncertainty of what will happen with the data and for how long it is kept*." An employee in the innovation and process management department in company D (D02, male, 52 years old) extended this thought and told us of a case that justified that type of concern: "*We had the case, where e-mails from the last 15 years were opened. I was also often involved. [It creates] fear that you are completely transparent.*"

4.4 Cyberbullying

Experiences with mobbing or bullying via ICT were the least prevalent type of stressor, comparable to techno-insecurity (A: 13%; B: 38%; C: 0%; D: 0%; AVG: 13%). In two companies, none of the interviewed individuals reported about any own experiences with stressors of this category, though they were mostly aware of the possibility of this type of behavior.

In those few cases were individuals reported on related experiences, they mostly traced rude or unsocial behavior back to the type of impersonal communication that ICT allow for. For example, a member of the management of company B (B05, female, 49 years old) felt that communication amongst employees was more confrontational if it was done via ICT: "*In our WhatsApp group we are often confronted with subservient messages by our drivers. What really annoys me is when pictures of a dirty vehicle get posted, just because there are some crumbs in it ... They make a mountain out of a molehill ... and everybody can see it*."

A potential explanation for the low prevalence of this type of stressor are coping mechanisms that individuals employ to avoid such issues related to ICT-mediated communication. One internal support employee in company A (A06, male, 23 years old) even stated that he now detests online communication and he explained that he tries to avoid it: "*I personally prefer to not write via ICT ... because something always comes back. In that moment I feel like [telling that person] that I am also not* *the perfect human being and also cannot deliver a 100% perfect answer. I am not interested in getting rebuked ten times in ten e-mails ….*"

4.5 Additional Technostressors

In addition to our eight stress categories that were defined and presented to the interviewed individuals, we also gave them the chance to talk about additional instances were ICT may lead to stressful experiences. With the exception of company B, individuals in every company told us about such instances (i.e., 13% of individuals in company A, 21% in company C, 20% in company D). We clustered these instances in two broader categories, with the first one being related to ambiguity in communication via ICT and the second one being related to information system security.

In company A and company C, some individuals expressed that they were sometimes confronted with confusion and misunderstandings when they communicated via ICT, about issues that could otherwise be resolved quickly when they interacted personally. For example, a customer support employee in company A (A07, male, 26 years old) reported about a case were a program misbehaved: *"[The function of the system] does not work and [another employee] shows it to me today and tells me that it does not work. I try all the necessary steps and it works. ... None of the software developers were able to [solve it]. They had exchanged 20 e-mails amongst each other, sometimes with 10 people in carbon copy ... and we solved the problem quickly in a casual interaction.*"

Relatedly, previous research has found that communicating the way we would usually do (e.g., using gestures and voice tonality) is mostly not possible using communication channels such as e-mail, which can lead to messages being interpreted in a different way [25].

The second area that was prevalent in companies C and D was stress from security threats, either directly (e.g., hacking attacks) or indirectly (e.g., stress from security measures to avoid security threats). A member of the sales department in company C (C04, female, 48 years old) reported on her concerns regarding the possibility of threats from outside the company: "*My brother-in-law told me that it is enough to open an e-mail. That can be enough for hackers to cause some damage ... and that is why you have to be really careful. For example, when an e-mail only with an attachment and nothing else comes in, then I am instantly nervous, because I do not know what it is. I immediately delete it then!*"

The steps that are then taken to avoid such threats should be planned carefully. If security measures are too rigorous, they might create stress on their own as indicated by a member of the marketing department in company D (D03, male, 48 years old): "*It becomes a problem when I want to load a website of one of our customers or potential customers and they include some words that are usually under embargo for us and ... the pop-up-blocker stops me ... that can be annoying right then*."

5 Conclusion and Future Research

Recently, Tarafdar et al. [26] presented updated definitions of their original constructs (see [1]), which now envelop some of our proposed stress categories. For example, aspects of unreliability are now included in techno-complexity as "interruptions, complications, hassles", monitoring is included as part of techno-invasion and security-related stress is spread out across techno-overload (adhere to security requirements), techno-invasion (surveillance and monitoring), techno-uncertainty (no control over IS use policies), and techno-complexity (hard to understand IS use policies). Yet, although the authors have acknowledged through this update that their original conceptualization of technostress was out-of-date to some degree ten years after its inception, or eventually lacking in completeness at all, we are not aware of an updated technostress measurement instrument that reflects this extended conceptualization on the item level.

In addition, it is questionable whether new aspects of technostress can simply be assigned to dimensions in the existing framework of Ragu-Nathan et al. [1] or whether additional dimensions are needed. For example, Ayyagari et al. [2] made a clear distinction between complexity and reliability as technology features. In our investigation, unreliability was the most prevalent technostress creator, but only occurred together with complexity to a limited extent (i.e., 48% of individuals who reported experiences with unreliability also had stressful experiences with complexity, which is less than uncertainty with 50% or overload with 63%).

Based on this evidence, we therefore make a call for additional quantitative investigations into the dimensionality of technostress, which acknowledges the current status of affairs. Based on the usage of a new set of items, it would be possible to demonstrate whether the same dimensions of technostress creators emerge again and therefore uphold the test of time or whether extending their definitions has led to a bloating of these first-order constructs beyond their conceptual usefulness.

Our research study also has limitations. First, though we used a sampling procedure that was comparable to Ragu-Nathan et al. [1] (i.e., firm-specific samples, selection of organizations based on contacts, self-selection of respondents), this naturally leads to a sample that is not generalizable to a wider population and results should therefore only be interpreted in the context of the specific companies and individuals involved. We also used a conservative approach in our coding procedure, only focusing on own accounts of previous stressful experiences. This may have led to an underestimation of certain stressor categories, for example indicated by the account of an employee in the research and development department of company D (D04, male, 30 years old): *"...when somebody is sitting next to you and receives a WhatsApp message or an SMS every 10 seconds and it rings all the time with some weird noise. That is not only stressful for them ... but it also stresses me if it rings all the time*." In addition, personal interviews may have led to some individuals not wanting to talk about their previous experiences in detail, as was sometimes the case for cyberbullying, where individuals hinted on some type of experience, but did not go into detail on their own (a form of social desirability bias). To minimize the influence of this bias, interviews were always conducted by a person that the individual was familiar with (to ensure a level of trust) using a standardized interview guide. We avoided conducting interviews ourselves as outsiders might be perceived as "agents" of the company management.

In summary, we argue that a systematic investigation of the factorial structure of technostress is needed. Further, we recommend the use of data triangulation in order to elicit stressor categories that were not considered before (e.g., bullying), but may be more important than self-report data alone could reveal initially.

6 Acknowledgements

This work was in part funded by the government of Upper Austria as part of the “Basisfinanzierung” funding initiative by the University of Applied Sciences Upper Austria, project title: “Digitaler Stress in Unternehmen”. This research was also funded by the Upper Austrian government as part of the PhD program “Digital Business International”, a joint initiative between the University of Applied Sciences Upper Austria and the University of Linz.

References

1. Ragu-Nathan, T.S., Tarafdar, M., Ragu-Nathan, B.S., Tu, Q.: The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. Information Systems Research 19, 417–433 (2008)
2. Ayyagari, R., Grover, V., Purvis, R.: Technostress: Technological Antecedents and Implications. MIS Quarterly 35, 831–858 (2011)
3. Maier, C., Laumer, S., Weinert, C., Weitzel, T.: The Effects of Technostress and Switching Stress on Discontinued Use of Social Networking Services: A Study of Facebook Use. Information Systems Journal 25, 275–308 (2015)
4. Gaudioso, F., Turel, O., Galimberti, C.: The mediating roles of strain facets and coping strategies in translating techno-stressors into adverse job outcomes. Computers in Human Behavior 69, 189–196 (2017)
5. Fischer, T., Riedl, R.: Technostress Research: A Nurturing Ground for Measurement Pluralism? Communications of the Association for Information Systems 40, 375-401 (2017)
6. Tarafdar, M., Tu, Q., Ragu-Nathan, T.S., Ragu-Nathan, B.S.: Crossing to the Dark Side: Examining Creators, Outcomes, and Inhibitors of Technostress. Communications of the ACM 54, 113–120 (2011)
7. MacKenzie, Podsakoff: Construct Measurement and Validation Procedures in MIS and Behavioral Research. Integrating New and Existing Techniques. MIS Quarterly 35, 293–334 (2011)
8. Tarafdar, M., Pullins, E.B., Ragu-Nathan, T.S.: Technostress: Negative Effect on Performance and Possible Mitigations. Information Systems Journal 25, 103–132 (2015)
9. D'Arcy, J., Herath, T., Shoss, M.K.: Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective. Journal of Management Information Systems 31, 285–318 (2014)

10. Shu, Q., Tu, Q., Wang, K.: The Impact of Computer Self-Efficacy and Technology Dependence on Computer-Related Technostress: A Social Cognitive Theory Perspective. International Journal of Human-Computer Interaction 27, 923–939 (2011)
11. Hudiburg, R.A.: Psychology of Computer Use. XXXIV. The Computer Hassles Scale: Subscales, Norms, and Reliability. Psychological Reports 77, 779–782 (1995)
12. Hudiburg, R.A.: Psychology of Computer Use. XVII. The Computer Technology Hassles Scale: Revision, Reliability, and some Correlates. Psychological Reports 65, 1387–1394 (1989)
13. Butler, B.S., Gray, P.H.: Reliability, Mindfulness, and Information Systems. MIS Quarterly 30, 211 (2006)
14. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress from a Neurobiological Perspective - System Breakdown Increases the Stress Hormone Cortisol in Computer Users. Business & Information Systems Engineering 4, 61–69 (2012)
15. Smith, M.J., Carayon, P., Sanders, K.J., Lim, S.-Y., LeGrande, D.: Employee stress and health complaints in jobs with and without electronic performance monitoring. Applied Ergonomics 23, 17–27 (1992)
16. Lohr, S.: Unblinking Eyes Track Employees. New York Times (2014)
17. Farmer, B.: More employees under surveillance at work. The Telegraph (2008)
18. Snyman, R., Loh, J.: Cyberbullying at work. The mediating role of optimism between cyberbullying and job outcomes. Computers in Human Behavior 53, 161–168 (2015)
19. Mayring, P., Fenzl, T.: Qualitative Inhaltsanalyse. In: Baur, N., Blasius, J. (eds.) Handbuch Methoden der empirischen Sozialforschung, pp. 543–556. Springer Fachmedien Wiesbaden, Wiesbaden (2014)
20. Myers, M.D., Newman, M.: The qualitative interview in IS research: Examining the craft. Information and Organization 17, 2–26 (2007)
21. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46 (1960)
22. McHugh, M.L.: Interrater reliability: The kappa statistic. Biochemia medica 22, 276–282 (2012)
23. Galluch, P., Grover, V., Thatcher, J.B.: Interrupting the Workplace: Examining Stressors in an Information Technology Context. Journal of the Association for Information Systems 16, 1–47 (2015)
24. Riedl, R., Fischer, T.: System Response Time as a Stressor in a Digital World: Literature Review and Theoretical Model. In: Nah, F.F.-H., Xiao, B.S. (eds.) HCI in business, government and organizations, 175-186. Springer Berlin Heidelberg, New York NY (2018)
25. Byron, K.: Carrying too Heavy a Load? The Communication and Miscommunication of Emotion by Email. Academy of Management Review 33, 309–327 (2008)
26. Tarafdar, M., Cooper, C.L., Stich, J.-F.: The technostress trifecta - techno eustress, techno distress and design. Theoretical directions and an agenda for research. Information Systems Journal 9, 1–37 (2017)

Application of Media Synchronicity Theory to Creative Tasks in Virtual Teams Using the Example of Design Thinking

Lukas Furmanek[1], Stephan Daurer[2]

[1]SAP Deutschland SE & Co. KG, Digital Business Services, Walldorf, Germany
lukas.furmanek@sap.com
[2]DHBW Ravensburg, Baden-Württemberg Cooperative State University, Department for Business Information Systems, Ravensburg, Germany
daurer@daurer.net

Abstract. Today in many industries it is common to work in virtual teams. At the same time companies must innovate to retain their market share and stay in the market. One possible method to come up with innovations is Design Thinking. However, this method is originally not intended to be used in a virtual team. The objective of this research is to provide a guideline on how to practice Design Thinking in a remote setup, thus travel expenses and planning efforts could be saved. This guideline is based on the media synchronicity theory which is applied to the process of Design Thinking. As a result, it is deduced that the diverging phases in Design Thinking are best supported with asynchronous media. In contrast to the converging phases which are supported best by synchronous media. These results could be used by practitioners to choose the best fitting media for each phase.

Keywords: Virtual Teams, Design Thinking, Media Synchronicity Theory, Creativity, Innovation

1 Introduction

To survive in a globalized market, companies have to come up with innovative services, new products and efficient processes [1]. Through constant innovation they can serve their customers better than the competition and even grow their market share. In order to innovate at scale, many companies started to work with creativity methods like Design Thinking to enhance their ability to come up with strong ideas [2].

At the same time, not only the markets became global markets, also the companies themselves opened locations all over the globe. For example, in the software industry, it is very common that teams are distributed over multiple locations and only communicate through information and communication technology. Such teams are called virtual teams. To communicate, they use, for example, e-mail, social networks, instant messaging, telephone conferencing and many more technologies. Nevertheless, also

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

virtual teams must solve creative tasks, to come up with innovations in their area of work.

This paper examines the intersection between these two fields. Therefore, we consider the usage of communication technology in virtual teams to solve creative tasks. As a method to solve creative tasks this paper employees Design Thinking, which enables a team to come up with innovative ideas. However, Design Thinking is originally a method, which is practiced in one location and not in a distributed set up. In this paper we develop a guideline on how a remote setup for virtual teams might look like. We apply the media synchronicity theory (MST), to analyze whether a certain medium fits to a task. With the perspective of the media synchronicity theory on Design Thinking, we focus on the research question "Which media capabilities should a tool for virtual Design Thinking have to enable high team performance?" The remainder of the paper is structured as follows. In the next section, we describe our methodological approach. In chapter three we explain how Design Thinking may be used to leverage creativity in teams and in chapter four we introduce the media synchronicity theory. After that we apply MST to the method of Design Thinking and we deduce recommendations for virtual teams. We conclude the paper with a discussion of our results and limitations.

2 Research Method

In this paper we apply a methodology from the domain of design science research [3–6]. We focus on the development of knowledge in information systems through the design of artifacts [7]. In our effort, we follow the guidelines for design science research as put forward by Hevner et al. [4]:

Design as an Artifact. The artifact that we develop in this paper is a concept that consists of guidelines for the application of specific types of media in virtual Design Thinking. The guidelines describe a category of media for every phase of Design Thinking which should enable a high task performance.

Problem Relevance. It is vital that research satisfies a strong business need. The business need for virtual Design Thinking originates from the conflict area between virtual teams, innovation needs and cost restrictions. As indicated above, Design Thinking can be used to come up with innovative solutions, however it is usually practiced in one place. In the case of a team with distributed team members, one solution is to travel to one place and to conduct the Design Thinking workshop there. This poses a large planning effort and it is rather expensive. This conflict between creativity and expenditure might be improved through the concept which we propose.

Design Evaluation. The artifact will be developed through the analysis and application of an existing theory (namely MST). We apply a deductive-argumentative approach using abstraction and reflection [8] when we relate MST to the different phases of Design Thinking.

Research Contribution. Our concept for virtual Design Thinking represents a first approach to make it possible to use Design Thinking in virtual teams. To the best of our knowledge, there is currently no guidance available for practitioners to use Design Thinking remotely.

Research Rigor. The contribution builds on the knowledge base in virtual team research, and Design Thinking. Therefore, a take a closer look on MST, and Design Thinking. Our approach is aligned with the research category of "design and action" which was proposed by Gregor [5].

Design as a Search Process. The starting point is the relevant practical problem, to which we apply a well-documented theory. The proposed concept is deduced from the application of MST to the Design Thinking process. In doing so, we analyze which capabilities are required to support each phase of the Design Thinking process. Hence, we are able to identify appropriate media for each phase.

Communication of Research. To communicate our research, we have designed this paper in a way that it addresses technology-oriented as well as management-oriented researchers. Furthermore, it is suited as a guideline for practitioners.

3 Leveraging Creativity in Teams Through Design Thinking

In the last decade, Design Thinking has evolved into a method, not only to support innovation, but also to offer a process to improve and to accelerate the creativity of teams [9]. Thus, Design Thinking isn't limited to the design space anymore. It can be used in different fields and for different applications. As it has evolved over time, multiple definitions of Design Thinking were developed [10]. Here we use the commonly accepted definition of Tim Brown (CEO of the innovation consulting firm IDEO): "Design Thinking is a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success." [11] This definition highlights that Design Thinking should aim at the intersection between the human desirability, the technical feasibility and business viability. If all these factors are given, an idea has the potential to be a break through innovation [12].

After a first glance on the purpose of Design Thinking, we describe the process. This process is iterative, so it is possible to cycle back, or to revisit the results of earlier phases and improve them again [10]. We use the Stanford Design Thinking Model (see Fig. 1), because that model allows to classify the phases according to the communication processes of the MST. In other models the phases are starting and ending at different times. That would make a matching to MST more difficult.

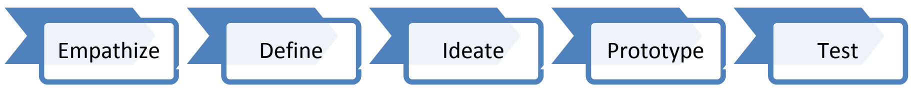


Figure 1: Stanford Design Thinking Model [13]

Empathize. The goal of this phase is to explore the problem and to get a common understanding [14]. An essential part of this phase is the knowledge transfer between the team members. Design Thinking teams should consist of experts from different areas to provide diverse inputs and to ensure a good result [15]. After the team members developed a common understanding of the problem and its context, the user research may be started. The goal of the user research is to develop empathy for the user. Therefore, the behaviors and feelings in the problem context have to be observed [13].

Define. The *definition* phase starts with the end of the phase *empathize*. The goal of this phase is to specify the problem and the context more precisely [14]. Furthermore, the user for whom the problem should be solved, will be defined. In a discussion about the results of the phase *empathize*, the team can find patterns in the data and also detect anomalies. Thus, the data can be framed into a narrow problem description [13].

Ideate. The *ideation* phase is the core of Design Thinking, in this phase most of the creativity and innovation takes place. In the first two phases, the empathy for the user is developed to guide the idea generation process in a good direction. Primarily, group brainstorming is used to generate new ideas in this phase [16]. However, also other smaller creativity methods are used in this phase.

Prototype. The *prototype* phase has two main goals. One is to concretize the generated ideas and the second is to make it possible to validate the concept with the user in the testing phase [13]. The prototype is helping the team and the customer to better understand the generated ideas [14]. Hence, possible constraints or limitations might be found in the testing phase and they might be used to improve the concept in a next iteration.

Test. The *test* phase is all about feedback from users and stakeholders. The ideas are presented to them with the help of the prototype. This allows a precise feedback [13]. The team will observe the customers' behavior while interacting with the prototype and ask them for feedback. Based on the feedback they are able to decide if the concept is ready for implementation yet.

4 Media Synchronicity Theory

In reviewing the literature about the media selection of virtual teams, we could find seminal relevant theories. According to the media richness theory (MRT), richer media should be used for tasks with high equivocality and simpler media should be used for standard tasks [17]. However, research has found that MRT does not explain the media choice well enough, because there are more factors that influence the media choice and different subtasks might have different media requirements [18]. The second theory considered was the task/technology fit theory. This theory explores the fit between a task and a medium on an individual [19] and group level [20]. The theory defines five different task types, which are supported by three different dimensions of media. This was an improvement on the issue of subtasks, however the research doesn't address if experience within the group might have an influence on the performance. MST tries to overcome these limitations and proposes the necessity of a fit between the communication processes, the appropriation factors and the capability of the medium to support synchronicity [21].

In MST there are two kinds of communication processes. Conveyance processes transport new and diverse information, so the receiver needs time to process and understand them [22]. In contrast, convergence processes transmit already known, or familiar information and the goal is to negotiate a shared understanding. MST proposes that conveying communication needs low media synchronicity and converging communication needs high media synchronicity [23]. Nevertheless, as shown in the model (see Fig. 2) also the appropriation factors, like familiarity with the medium and past collaborations between the team members have influence on the media fit. In the next sections, all factors will be discussed in more detail, so a good understanding of the theory can be developed.

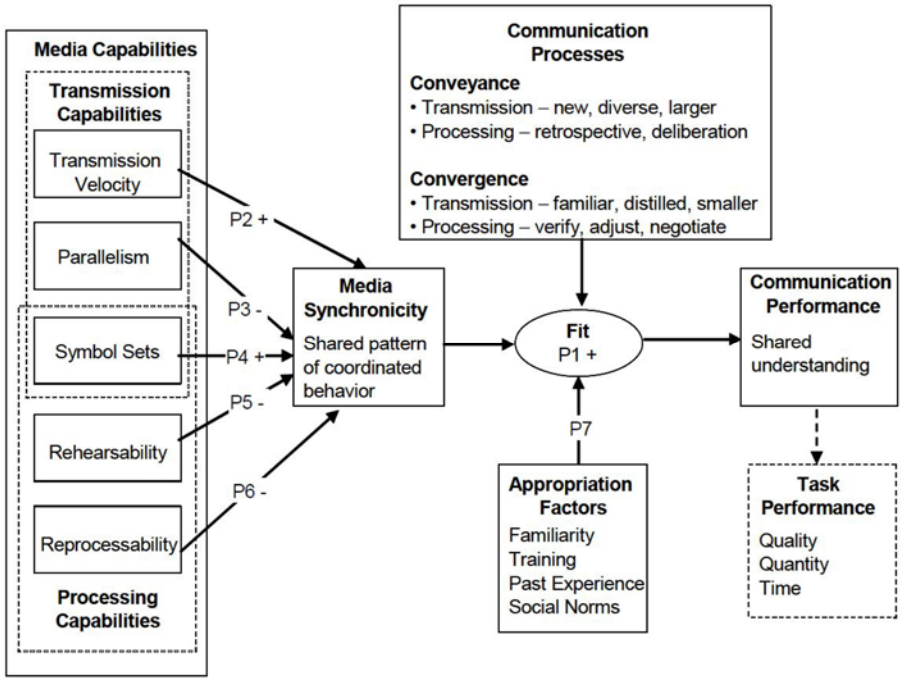


Figure 2: Media Synchronicity Theory [22]

4.1 Communication Processes

There are two types of communication processes. First, conveyance processes are used to deliver a large amount of new and diverse information to the receiver. This might be information like a report, but also a collection of raw information. So the receiver has to process the information and to interpret the meaning [24]. To fulfill this task the receiver needs time to go through the heap of information. For the receiver this task gets easier, if he has the possibility to revisit the content of the message multiple times and to make sense out of it [25]. On the other hand, the sender could support the receiver by carefully crafting the messages [26].

Second, in contrast, converging communication deals with already familiar information or preprocessed information [22]. In this case the participants of the conversation already have given an own meaning to the information. Hence, the goal is to discuss towards a common understanding of the information [27]. Therefore, only small quantities of information are transmitted. The receiver doesn't need much time to process the information. A high transmission velocity could ensure that sender and receiver have a smooth and fast communication [28].

Almost every task involves a mix between the two communication types. Thus, the theory can be applied to the dominant part of the communication processes. Or multiple media types can be combined to support the communication best [22]. Larger tasks like the development of a concept by using Design Thinking can be divided into subtasks. So, the subtasks might be analyzed, and appropriate media can be used.

4.2 Media Capabilities

In this section we describe the media capabilities which define the media synchronicity. A medium with high media synchronicity, like video conferencing, allows its participants to communicate almost synchronously [29]. In contrast, there are media with low synchronicity, for example, e-mail. According to MST there are five media capabilities, which build the medium's capability to support synchronicity. These capabilities are transmission velocity, parallelism, symbol set, rehearsability and reprocessability [23]. The first three media capabilities are adopted from the Shannon and Weaver communication theory [30]. Rehearsablity and reprocessability were added to describe the influence of encoding and decoding in the communication theory [22].

Transmission velocity describes, if the medium supports a fast feedback on the understanding and the meaning of the last message [25]. As indicated by the plus symbol next to P2 in figure 2 a high transmission velocity causes a higher media synchronicity. The opposite is the case for parallelism here the minus symbol indicates that high parallelism causes lower media synchronicity.

Parallelism describes the capability of a medium to support multiple conversations at a time [26]. The media synchronicity is lower, because the participants aren't focusing on one conversation.

Symbol set describes the mediums capabilities to encode a message [28]. For example, in phone calls there is only sound to transport the information, but video conferencing additionally has the video and as a consequence a larger symbol set.

Rehearsability describes the mediums capability to rephrase the message before sending it [31]. A high rehearsability indicates a low media synchronicity.

Reprocessability describes the mediums capabilities to replay or show the received message multiple times [22]. So, the receiver can reread the message while making sense out of it in conveyance processes.

Table 1 sums up, which media supports high synchronicity and which media supports low synchronicity.

4.3 Appropriation Factors

The appropriation factors are the third part which has to fit to ensure a high communication performance. The appropriation factors describe whether the group already has experience working together, working on a similar task, or working with a certain communication medium [22]. The idea behind the appropriation factors is: If the team has no experience with one, or more aspects out of *people*, *task* and *media* there is more coordination required [21]. Consequently, a higher media synchronicity is required, since coordination mainly involves convergence processes. Nevertheless, over time the team develops a shared understanding of the task, the media and the people on the team [22]. As a result, the need for synchronous media out of the appropriation factors decreases over time.

Table 1: Classification of various media regarding MST's media capabilities [31]

Media	Immediacy of feedback	Symbol variety	Parallelism	Rehearsability	Reprocessability
Face-to-face	High	High	Low	Low	Low
Video conference	Med.-High	Medium	Low	Low	Low
Telephone conference	High	Medium	Low	Low	Low
Instant messaging	Med.-High	Low-Med.	Low-Med.	Medium	Med.-High
Synchronous electronic conferencing	Med.-High	Low-Med.	Low-Med.	Low-Med.	Medium
Asynchronous bulletin board	Low-Med.	Low-Med.	High	High	High
Asynchronous e-mail	Low-Med.	Low-Med.	High	High	High

5 Application of Media Synchronicity Theory on Design Thinking

In this chapter MST is applied to the Design Thinking process. However, Design Thinking involves both conveyance and convergence communication processes. In order to get a more specific result, the Design Thinking process phases will be classified into categories. As discussed by Lindenberg et al., there are diverging and converging thinking phases. Figure 4 shows, that the *empathize* and the *ideation* phase are divergent phases and the phases *define*, *prototype* and *test* are convergent phases [32, 33]. The small circles represent the ideas which are created in the diverging phases and afterward are chosen by the group to work on further, or not. For example, in the phase *empathize* the team explores multiple problems and user groups, which are involved in the design challenge. In this stage, they create choices, by exploring many options. In the following definition phase, the team must take the choices and for example decide for a user group. As the example shows at first the team have to think diverse, to explore all the possibilities and afterwards they have to converge to a group point of view [34]. For the ideation, followed by prototyping and test, the argumentation is analogous. At first, a large amount of ideas is created and then the team selects the best one for a prototype.

In a kickoff meeting before the *empathize* phase the team has the chance to get a first impression of the *task*, the *people* and the *media*. In most situations one of the factors is new, even when the team is used to work in a remote setup. However, in many cases either an entirely new team is formed, or the group is complimented with experts from different areas and Design Thinking coaches. The kickoff meeting should be used to discuss the process, the media which should support every phase and to introduce the team members. The meeting should give the team the chance to discuss questions, get to know each other and negotiate the working mode. All these tasks involve convergent communication processes and are therefore best supported by synchronous media.

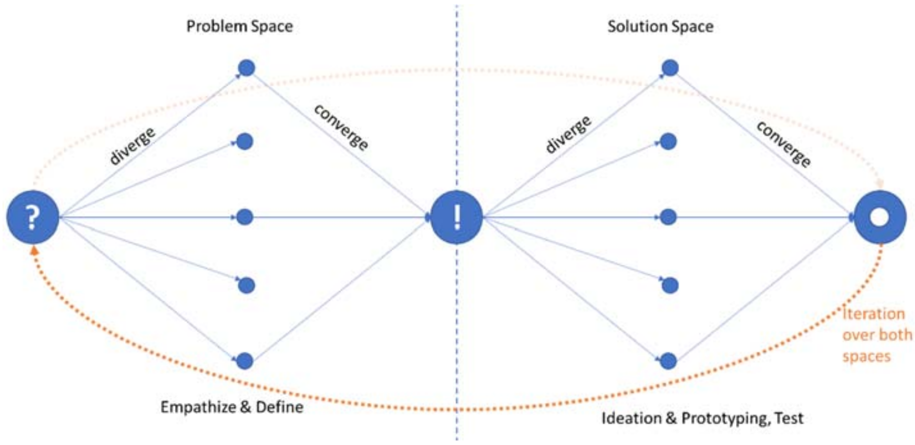


Figure 3: Problem and solution space in Design Thinking adopted to the Stanford Design Thinking model [32, 33]

5.1 Diverging Phases

As described in the section on Design Thinking, in the phase *empathize* the different experts sharing their views on the problem space. Additionally, the user research takes place. These two exercises mainly involve conveyance communication, because a large part of it is sharing the different points of views. For this the team could use a digital whiteboard, which allows to work in a similar mode as in an onsite workshop. In the virtual setup the team shouldn’t only capture the thought in a few words, but also add some explanation to make it easier for others make sense out of it. Nevertheless, there are also some convergence processes involved. Thus, it could be beneficial to do a video conference before the user research takes place. This enables the team to discuss a common questionnaire and work on the appropriation factors. Afterwards each team members can interview people at their location and convey the information back to the team. Consequently, there are synchronous and asynchronous media capabilities required in the phase *empathize*, with a tendency to asynchronous media.

In the *ideation* phase, the main goal is to create as many ideas as possible. In this phase, all participants share their ideas and build new ideas on top of ideas from other participants. The shared information is mainly new and diverse information, which should be considered by the other participants. According to the MST these are the characteristics of conveyance communication, resulting in a bigger need for asynchronous media. By using this type of media, the team could make use of the high rehearsability and reprocessability. The team would carefully craft the messages so that the ideas are easy to understand. In addition, there would be always the possibility to revisit an earlier message.

When there are multiple time zones involved, the divergence phases *ideation* and *empathize*, also allow to work at different times. Thereby, the duration of the work will increase due to less parallel work. The working time on the tasks itself should be comparable to onsite workshops. However, in the first phases the team might need some more time to find a proper working mode.

5.2 Converging Phases

After multiple problem areas and user groups were explored in the phase *empathize*, the team has to specify a smaller subset in the *definition* phase to go on with. Such decisions require a lot of coordination and interactions between the team members [22]. They involve a quick back and forth between the team members, as they discuss which problem should be prioritized. This dominant part of the conversation is optimally supported by synchronous media, since it involves mainly convergence communication. Synchronous media provides the team with a high immediacy of feedback and a rich symbol variety. For example, by using a video conferencing each team member becomes fast feedback on the arguments used and will observe additional facial or gestural reactions. However, it can be argued, that a combination with a form of asynchronous media could be also helpful in this stage. An asynchronous medium, like a digital whiteboard, could support the team in this stage by providing an overview on the results of the last session and give the group guidance in this session.

Similar to the *definition* phase also in *prototyping* and *test* the team has to make some decisions. Furthermore, the group has to build the prototype and discuss the results in the testing step. Prior to building the prototype the group has to select the best solution to prototype. Analogously to the *definition* phase this requires a lot of discussion and coordination, which makes it a converging communication. On the other hand, the team also has to build a prototype, which could involve more inputs from one of the experts. So again, a mix between synchronous and asynchronous media is needed.

5.3 Media Capabilities for a Tool to Support Virtual Design Thinking

This section summarizes the results of the analysis and give some guidance, which media types might be used to reach a good team performance in virtual Design Thinking. Summarized in one sentence: Asynchronous media are appropriate for the diverging thinking phases and synchronous media are better suited for the converging thinking phases.

More precisely, the start of a virtual Design Thinking workshop is best supported by a synchronous medium. It supports the participants best in getting to know each other and to get a common understanding of the way of collaborating. Therefore, a medium like video conferencing, which supports a high immediacy of feedback and offers a high symbol variety, would be appropriate. As the team begins to work on the first task (*empathize*), a more asynchronous medium will be appropriate, like a digital whiteboard (cp. the asynchronous bulletin board in table 1). This allows the team to profit from high parallelism, rehearsability and reprocessability. As described in Redlich et al. [35] a digital whiteboard and a chat could be used permanently. With that combination the distributed Design Thinking teams almost reached the performance of the onside Design Thinking teams [35].

In the *definition* phase, the team could be using video conferencing again to profit from the synchronicity of the medium. At the same time the whiteboard could be used for documenting the results and facilitate the session. Afterwards, in the *ideation* phase the team goes back to using an asynchronous medium. Finally, the team will reach the *prototyping* and *testing* phase, which again requires a synchronous media, like telephone conferencing.

6 Conclusion and Outlook

In Design Thinking there are two main styles of thinking, the converging thinking and the diverging thinking. Hence, the research question "Which media capabilities should a tool for virtual Design Thinking have to enable high team performance?" had to be answered for both thinking styles. For phases, which involve diverging thinking a medium with low synchronicity will lead to a higher team performance. And on the other hand, for phases which involve converging thinking synchronous media should be preferred. A medium which supports high synchronicity has the following media capabilities: It has high transmission velocity, low parallelism, a large symbol set, low rehearsability and low reprocessability. Media which support low synchronicity have the opposite media capabilities. By transferring the media synchronicity theory to the domain of Design Thinking we propose recommendations of which media type is suited in different phases of virtual Design Thinking. Therefore, we were able to answer the initial research question in an argumentative and deductive approach. However, as any research, this study comes with some limitations: So far, there is no empirical evidence for the results yet. To the best of our knowledge Redlich et al. [35] is currently the only study, which indicates that virtual Design Thinking might be successful. Yet, this study used a digital whiteboard in combination with a messenger. In contrast, the guideline proposed in this paper also employs synchronous media in convergence communication phases, which should increase the task performance further. To validate the approach of our work, future research could validate the general approach of our work in case studies. In addition, a bigger empirical evaluation with different team setups involving multiple time zones could be done to provide further evidence.

References

1. Gabriel, A., Monticolo, D., Camargo, M., Bourgault, M.: Creativity Support Systems. A Systematic Mapping Study. Thinking Skills and Creativity 21, 109–122 (2016)
2. Redlich, B., Siemon, D., Lattemann, C., Robra-Bissantz, S.: Shared Mental Models in Creative Virtual Teamwork. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 464–473 (2017)
3. March, S.T., Gerald F. Smith: Design and Natural Science Research on Information Technology. Decision Support Systems 15, 251–266 (1994)
4. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–106 (2004)
5. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
6. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–78 (2007)
7. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37, 337–355 (2013)
8. Gregor, S., Müller, O., Seidel, S.: Reflection, Abstraction And Theorizing In Design And Development Research. In: Proceedings of the 21st European Conference on Information Systems, pp. 1–12 (2013)
9. Tschimmel, K.: Design Thinking as an effective Toolkit for Innovation. In: ISPIM Conference Proceedings, pp. 1–20 (2012)
10. Liedtka, J.: Perspective. Linking Design Thinking with Innovation Outcomes through Cognitive Bias Reduction. Journal of Product Innovation Management 32, 925–938 (2015)
11. Brown, T.: Design Thinking. Design Thinking is a Process for Creative Problem Solving, https://www.ideou.com/pages/design-thinking
12. Weiss, L.: Developing Tangible Strategies. Design Management Review 13, 33–38 (2002)
13. Henriksen, D., Richardson, C., Mehta, R.: Design Thinking. A Creative Approach to Educational Problems of Practice. Thinking Skills and Creativity 26, 140–153 (2017)
14. Grots, A., Pratschke, M.: Design Thinking. Kreativität als Methode. Marketing Review St. Gallen 26, 18–23 (2009)
15. Brown, T., Wyatt, J.: Design Thinking for Social Innovation. Stanford Social Innovation Review 12, 29–43 (2010)
16. Seidel, V.P., Fixson, S.K.: Adopting Design Thinking in Novice Multidisciplinary Teams. The Application and Limits of Design Methods and Reflexive Practices. Journal of Product Innovation Management 30, 19–33 (2013)
17. Daft, R.L., Lengel, R.H., Trevino, L.K.: Message Equivocality, Media Selection, and Manager Performance. MIS Quarterly 11, 355–366 (1987)
18. Dennis, A.R., Valacich, J.S., Speier, C., Morris, M.G.: Beyond Media Richness. An Empirical Test of Media Synchronicity Theory. In: Proceedings of the 31st Hawaii International Conference on System Sciences, pp. 48–57 (1998)
19. Goodhue, D.L., Thompson, R.L.: Task-Technology Fit and Individual Performance. MIS Quarterly 19, 213–236 (1995)
20. Zigurs, I., Buckland, B.K.: A Theory of Task/Technology Fit and Group Support Systems Effectiveness. MIS Quarterly 22, 313–334 (1998)
21. Lam, C.: Improving Technical Communication Group Projects. Journal of Business and Technical Communication 30, 85–112 (2015)

22. Dennis, A.R., Fuller, R.M., Valacich, J.S.: Media, Tasks, and Communication Processes. A Theory of Media Synchronicity. MIS Quarterly 32, 575–600 (2008)
23. George, J.F., Carlson, J.R., Valacich, J.S.: Media Selection as a Strategic Component of Communication. MIS Quarterly 37, 1233–1251 (2013)
24. Robert, L.P., Dennis, A.R.: Paradox of Richness. A Cognitive Model of Media Choice. IEEE Transactions on Professional Communications 48, 10–21 (2005)
25. Niinimäki, T., Piri, A., Lassenius, C., Paasivaara, M.: Reflecting the Choice and Usage of Communication Tools in Global Software Development Projects with Media Synchronicity Theory. Journal of Software: Evolution and Process 24, 677–692 (2012)
26. Andres, H.P.: Team Cognition using Collaborative Technology. A Behavioral Analysis. Journal of Managerial Psychology 28, 38–54 (2013)
27. Maruping, L.M., Agarwal, R.: Managing Team Interpersonal Processes through Technology. A Task-Technology Fit Perspective. The Journal of Applied Psychology 89, 975–990 (2004)
28. Maynard, M.T., Gilson, L.L.: The Role of Shared Mental Model Development in Understanding Virtual Team Effectiveness. Group & Organization Management 39, 3–32 (2014)
29. Carlson, J.R., George, J.F.: Media Appropriateness in the Conduct and Discovery of Deceptive Communication. The Relative Influence of Richness and Synchronicity. Group Decision and Negotiation 13, 191–210 (2004)
30. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. The University of Illinois Press (1964)
31. Davison, R., DeLuca, D., Valacich, J.S.: Virtual Teams In and Out of Synchronicity. Information Technology & People 19, 323–344 (2006)
32. Lindberg, T., Meinel, C., Wagner, R.: Design Thinking. A Fruitful Concept for IT Development? In: Plattner, H., Meinel, C., Leifer, L.J. (eds.) Design Thinking. Understand - improve - apply, pp. 3–18. Springer, Heidelberg, London (2011)
33. Thoring, K., Müller, R.M.: Understanding the creative mechanisms of design thinking. an evolutionary approach. In: Martens, J.-B., Markopoulos, P. (eds.) Procedings of the Second Conference on Creativity and Innovation in Design, pp. 137–147. ACM Press, New York, New York, USA (2011)
34. Leifer, L.J., Steinert, M.: Dancing with Ambiguity. Causality Behavior, Design Thinking, and Triple-Loop-Learning. Information Knowledge Systems Management 10, 151–173 (2011)
35. Redlich, B., Dorawa, D., Siemon, D., Lattemann, C.: Towards Semi-Virtual Design Thinking. Creativity in Dispersed Multicultural and Multidisciplinary Innovation Project Teams. In: Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 717–726 (2018)

TrustyTweet: An Indicator-based Browser-Plugin to Assist Users in Dealing with Fake News on Twitter

Katrin Hartwig and Christian Reuter

Technische Universität Darmstadt,
Science and Technology for Peace and Security (PEASEC), Germany
katrin.hartwig@stud.tu-darmstadt.de
reuter@peasec.tu-darmstadt.de

Abstract. The importance of dealing with fake news on social media has increased both in political and social contexts. While existing studies focus mainly on how to detect and label fake news, approaches to assist users in making their own assessments are largely missing. This article presents a study on how Twitter-users' assessments can be supported by an indicator-based white-box approach. First, we gathered potential indicators for fake news that have proven to be promising in previous studies and that fit our idea of a white-box approach. Based on those indicators we then designed and implemented the browser-plugin *TrusyTweet*, which assists users on Twitter in assessing tweets by showing politically neutral and intuitive warnings without creating reactance. Finally, we suggest the findings of our evaluations with a total of 27 participants which lead to further design implications for approaches to assist users in dealing with fake news.

Keywords: Fake News, Social Media, Twitter, Plugin

1 Introduction

Fake news can be defined as "*news articles that are intentionally and verifiably false and could mislead readers*" [1]. Recently, the term has gained popularity, especially in discussions concerning the political context. The U.S. presidential election in 2016, as well as the German parliamentary election in 2017 among others, showed a great perceived significance of fake news for the society. Although studies have shown that there were no impacts on the election outcomes [1], the society fears the effect of fake news in social media. Our previous representative study on the perception of fake news in Germany revealed that 84 % of the citizens agree with fake news posing a threat [2]. Those concerns are not groundless as fake news can indeed have serious consequences. For example, in 2013 the official Twitter account of Associated Press (AP) was hacked. In consequence, the stocks experienced a temporary loss of $130 billion [3]. Furthermore, fake news can be relevant in the context of peace and political propaganda [4]. Thus, finding adequate strategies to counteract the negative effects of fake news, especially in social networks, is of high interest. Examining fake news in online information is highly relevant in the IS research field [5]. Several studies have already

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

shown that labeling and deleting fake contents is not effective and sometimes counterproductive. Instead, scientists argue that the training of media literacy is a promising strategy [6], [7]. (Media) literacy is defined as the ability to access, analyze, evaluate and create messages in a variety of forms [8]. However, most approaches concentrate on black-box algorithms to automatically detect and label fake news. In black-box approaches one can observe the input (in our case e.g. a tweet) and the output (here e.g. the label as "fake") but there is no information about what happens in between (e.g. why the tweet was labeled as "fake"). The counterpart is called white-box approach, where internals can be reviewed. In our context, white-box approaches facilitate the comprehension of reasons that indicate fake content, so that the user has all necessary information to understand why an algorithm has a specific output.

The objective of this article is to examine how users on Twitter could be supported in dealing with fake news by a white-box-based browser-plugin. Our research questions are: How can we provide a transparent, politically neutral and objective assisting tool for users of social media? Moreover, does a white-box approach counteract reactance and encourage a learning effect? The article is organized as follows: Section 2 presents related work on assisting tools to counteract fake news. Section 3 presents our research approach of *design science* [9], which focuses on the design of an artifact for a relevant problem and rigorous evaluation methods. In section 4 we propose the concept of *TrustyTweet*, a white-box plugin for users on Twitter whose evaluation will be presented in section 5. Finally, we discuss the potential scientific contributions and limitations of our approach in section 6.

2 Related Work & Research Gap

While the effect of fake news has proven to be significant in specific cases and the debates in politics and society continue, several approaches try to find answers on how to counteract fake news. Recently, many studies have been conducted to detect and label fake news. Viviani and Pasi present a survey on how approaches automatically assess credibility in online review sites, microblogs and sources of online health information [10]. Rubin presents the state-of-the-art technologies on fake news detection, divided into linguistic and network approaches [11]. Both studies show that most of the approaches use machine learning techniques to identify fake contents. Despite machine learning algorithms, blacklists and whitelists of websites are commonly used. In the following list, an excerpt of existing browser-plugins and smartphone apps is presented which we found searching the Google Play Store, browser add-on sections and scientific contributions. Relevant associated characteristics were extracted from the official descriptions and are given in **Table 1**.

- **TweetCred**: Browser-plugin with a semi-supervised ranking model using SVM-rank to assess credibility in tweets. Displays a credibility score from 1 to 7. [12]
- **B.S. Detector**: Searches all links on a given webpage for references to unreliable sources, checking against a manually compiled list of domains. (http://bsdetector.tech)

- **Fake News Detector**: Allows users to tag news stories on Facebook and Twitter. Tags will be stored in a database and used by an AI to learn. In the future contents will be highlighted based on user input and the AI. (https://fakenewsdetector.org)
- **Fake News AI**: Uses a neural network to analyze writing, sophistication, site popularity, content and many more. (http://www.fakenewsai.com)
- **Fake News Check**: A smartphone app that does not detect fake news automatically but causes to reflect by asking 19 relevant questions which the user needs to answer to receive feedback. The app was developed for students to train media literacy. (https://www.neue-wege-des-lernens.de/projekte/fake-news-check)

Table 1. Limitations of existing approaches

Name of application	*Binary labels*	*Gives transparent reasons*	*Provides learning effect*	*Uses a database (black- or whitelist)*	*Based on training data*	*Big effort for users*
TweetCred					X	
B.S. Detector	X			X		
Fake News Detector				X	X	X
Fake News AI	X				X	
Fake News Check		X	X			X

Rehm states that fully automatic technologies are partly suitable to support the user in dealing with fake news but cannot take over all necessary tasks. Additionally, approaches have to be based on human intelligence [13]. The stand-alone usage of blacklists and whitelists works only for websites or other texts that contain links to URLs included in one of those lists. Since the online environment is much more complex, manually compiling lists with reliable or unreliable sources does not lead to sufficient results [10]. Despite the absence of gold standard datasets to train the classifiers [10], machine learning algorithms have another major flaw if used to assist the user. As machine learning techniques are black-box procedures, they cannot reason why they label a content as fake. Showing the user a label, it might even create reactance if it does not fit his or her own perception. That effect is caused by the *Confirmation Bias* due to which messages are particularly considered true if they fit the own ideology [14–16]. None of the existing approaches gives transparent reasons or encourages a learning effect while leading to little effort for the user.

Studies have shown that a promising way to counteract fake news is increasing media literacy [17, 18]. Improving the capacity to evaluate online contents as autonomously as possible using white-box instead of black-box approaches can minimize reactance and prevent the *Backfire Effect*. While several approaches use machine learning techniques to label contents as fake or not fake, there are no approaches that work with a white-box technique despite “Fake News Check”. This application, though, involves a big effort for the user as he must manually go through 19 questions for every text he wants to check before receiving a feedback. The approach itself does not include an automatic check to detect fake news. To date, no approach to detect fake news is based on the ideas of media literacy or white-box methods.

3 Research Design

Keeping in mind the detected limitations of machine learning-based approaches, for instance regarding created reactance, our intention is to offer a first survey on how white-box approaches can help to counteract fake news in social media. Our aim is to find answers to the following research questions: *How can we provide a transparent, politically neutral and objective assisting tool for users of social media? Furthermore, does a white-box approach counteract reactance and encourage a learning effect?* To encounter the lack of empirical findings regarding white-box approaches in the given context, we present a browser-plugin for Twitter which has been developed and evaluated in an iterative process. The plugin focuses on Twitter as it is a popular platform for breaking news with a high relevance of fake news, for instance in emergency situations but also in political campaigns [19]. As a popular communication channel, it is commonly used in scientific studies to examine social media from various perspectives (e.g. [20 21]). We used the *design science* approach [9] which focuses on the design of an artifact for a relevant problem and rigorous evaluation methods. In our case, the artifacts will be versions of our plugin and evaluations will take place in form of thinking-aloud studies. The method applies five steps, namely (a) achieving a problem awareness, (b) suggesting solutions, (c) development of solutions, (d) evaluation and finally (e) conclusion. The *design science* approach has proven to be an appropriate method to create new and innovative artifacts [9].

4 Concept of *TrustyTweet*

Instead of labeling and deleting, acquiring a high standard of media literacy is considered to be a promising approach in combating the impact of fake news. Given a number of transparent and identifiable indicators for fake news, the user of social media can be supported in forming an opinion about online content. In that context, it is crucial to differentiate between assistance systems that give neutral hints based on transparent indicators to train media literacy and systems that create reactance. Müller and Denner indicate that warning messages might lead to a *Backfire Effect* [6]. Especially in political contexts, users might rate the warning message as an illegal attempt to persuade the user which can result in believing in the content even more. Using a white-box instead of a black-box procedure is an important step to prevent or minimize reactance.

Our aim is to support the user in dealing with fake news in tweets and to increase his media literacy. We present *TrustyTweet*, a browser-plugin that intends to support users on Twitter in dealing with fake news by giving politically neutral, transparent and intuitive hints. This approach particularly aims to be a helpful assistant without creating reactance. The user continues to be in the power of his own assessments. We intend to create a learning effect regarding media literacy to make the plugin redundant after a longer regular usage. Therefore, different from other approaches, *TrustyTweet* is based on white-box technology. The plugin was developed in a user-centered design process using the design science approach. We identified potential indicators for fake news by considering what has already been proven in studies to be successful. The focus is on heuristics that are used intuitively and successfully by humans and are easily comprehensible for everyone. In several qualitative thinking-aloud studies we evaluated the perceived helpfulness, the users' perceived autonomy and usability of the plugin on Twitter.

4.1 Identification of Indicators

Since the plugin aims to be a white-box approach it intends to show transparent warnings which the user can comprehend at any time, regardless of his level of media literacy. Morris et al. [22] found that users assess contents especially using features that are visible at a glance. In our approach, we intend to follow that idea. Potential indicators that fit our intentions and that largely have already been proven in studies to be promising in indicating fake news are the following:

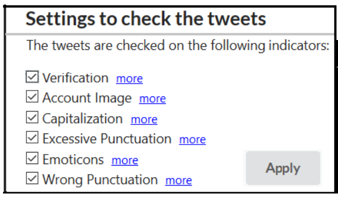
- Consecutive capitalization [23-25]
- Excessive usage of punctuation (e.g. "*!!!*") [22], [23]
- Wrong punctuation a the end of sentences [22], [24]
- Excessive usage of emoticons [23], [25] and attention-grabbing emoticons
- Default account image [22]
- The absence of an official account verification seal (especially for celebrities) [22]

The potential indicators were assessed in a first thinking-aloud pre-study with six participants aged 23 to 28 (4 female, 2 male, all university students) in an average duration of 33 minutes. All six indicators have proven to be easily comprehensible for our test subjects. Furthermore, the detection algorithm for each indicator has an acceptable runtime to support users on Twitter dynamically and in real time. It is vital to clarify that our approach does not claim to comprehend all relevant indicators for fake news. The mentioned group of indicators includes those that are fitting our white-box idea since they are easily comprehensible.

4.2 Underlying Technology and Components of the Plugin

TrustyTweet was developed for the Firefox browser and uses jQuery and Semantic UI. Its main components are a textbox containing all indicators detected in a specific tweet which serves as a warning, two distinct icons to report if indicators have been detected in the specific tweet or not and an icon to open the settings in a pop-up window. The indicators were detected by searching the DOM tree of Twitter. Next to each indicator, there is a link to open more generic information about the indicator in a pop-up window (see **Figure 3**). When hovering the mouse over an indicator, the underlying component of the tweet is being highlighted dynamically (see **Figure 1**). Hence, the user can see immediately why the warning is being displayed. The main icon of the plugin serves as a toggle button for the textbox. The user can decide if he wants to see all detected indicators next to each tweet or if he prefers to see only the icon and toggle the textbox whenever he is interested in why the warning is being displayed. Additionally, it guarantees that other contents like “*Who to follow*” do not remain hidden. A central feature of *TrustyTweet* is the configuration pop-up (see **Figure 2**). Using checkboxes, the user can switch the examination regarding specific indicators on and off. Hence, the plugin intends to build a stronger sense of autonomy and to counteract paternalism.

Figure 1. Exemplary output of plugin for four tweets



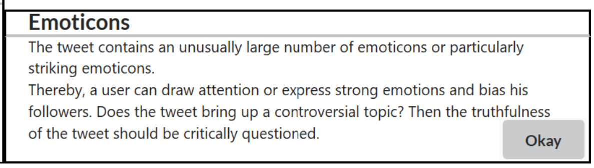


Figure 2. Pop-up with settings

Figure 3. Pop-up with additional information

5 Evaluation

5.1 Methodology

Using the design science approach, we iteratively applied five steps to achieve a problem awareness, to suggest solutions in form of potential plugin-designs, to implement those solutions, to thoroughly evaluate them and to finally draw a conclusion. The iteratively conducted evaluations were based on the thinking-aloud method in which the user explains why he carries out which activity, which information is incomprehensible or does not meet his expectations and what he likes or dislikes. The audio and video material was recorded using the Xbox DVR-tool. While using *TrustyTweet* on Twitter, the subjects were asked to execute usability tasks and answer open questions. The average durations were 33 minutes in the first pre-study and 11 minutes in the second and main study. The participants were informed beforehand that there is no "*right*" or "*wrong*" for answering the questions. To guarantee the same conditions for all subjects in the first subtest, tweets were generated by a test account by which the subtest was performed. Additionally, a second subtest was performed on real-time tweets of a politician and a German news page on Twitter to receive an impression of the usage of the plugin in a realistic environment. After each study, the detected flaws were patched and suggested improvements were implemented to receive better results in the next iteration. While the tasks in the first formative study focus mainly on the comprehensibility of the suggested indicators and gather ideas on how to increase the perceived autonomy, the second study includes an examination of the central configuration feature and the realization of dialogue design principles. Since interactions are a central component of our plugin, we want to gather information about the fulfillment of dialogue principles as conformity with user expectations and self-descriptiveness. The third study intends to examine to what extent a well-usable version of the plugin supports the user in dealing with fake news, including a summative evaluation of usability. Continuing the iterative process our aim is to understand to what extent users feel autonomous or patronized during the usage of our white-box-based plugin. Furthermore, we intend to examine the perceived helpfulness of the plugin. Therefore, the test subjects were asked to perform several tasks (e.g. Open the configurations of the plugin. Check the tweets on "*capitalization*" and "*emoticons*" only.) and answer specific questions (e.g. To what extent do you feel patronized or autonomous when assessing the tweets? / How do you like the plugin contentwise? Is it helpful or is it obstructive?).

Characteristics of Study Participants. In the first pre-study, a number of six participants (4 female, 2 male) took part, in the second pre-study a number of five (2 female, 3 male) and in the main study a number of 16 participants (7 female, 9 male). The participants' age ranged from 23 to 28 years in the first and second study and from 21 to 34 years in the main study. The majority of the test subjects were university students (19 out of 27 in total) due to their good accessibility for scientific studies and their relevance as potential Twitter-users. The remaining eight participants stated to be employees. In the first pre-study, three out of six participants and in the second pre-study three out of five participants stated that they have a Twitter account or that they had an account in the past. In the main study, it applied to half of the test subjects. Participants in all studies that stated to have never had a Twitter account were introduced to the central aspects and components of Twitter and its tweets before they started completing the tasks.

Analysis. Following the standard proceeding for thinking-aloud tests according to van Someren, Barnard & Sandberg [26] we examined the obtained qualitative data of all thinking-aloud studies by reviewing the video and audio material, transcribing all important statements and assigning the statements to their associated tasks and actions. The statements were then clustered thematically and gathered for all test subjects. Eight categories were developed inductively from the data (helpfulness, autonomy, additional information, configuration, toggle-feature, mouseover-feature, salience, other). Each statement was assigned to a category by looking for keywords (e.g. "patronized": autonomy), considering the context of tasks. Hence, conclusions were drawn from the various categories. The most noteworthy contributions of the main study will be presented in the following chapter.

5.2 Empirical Results

Perceived Helpfulness. When asked about helpfulness or obstructiveness of the plugin, 13 out of 16 participants regarded it to be a helpful tool. Most participants appreciated particularly its transparent nature and the simple visual feedback as well as the possibility to toggle the textbox while still getting a feedback from the icon: *"I like it a lot. You must keep thinking for yourself, but the plugin makes it easier and things attract your attention faster. It says: Attention! Here it would be wiser to think about the tweet again" (E12 #00:09:28).*

On the other hand, three participants argued that they personally do not need the plugin and therefore could not yet see the added value. They pointed out, that the warnings were based on very simple indicators, which they were able to detect by themselves: *"I think I do not need it. It is very interesting, that the displayed warnings are exactly the things I use as a search filter in my head when I read texts." (E15 #00:08:53).* One participant was concerned about the plugin showing too many false alarms, for example when warning of non-celebrity users that are not verified: *"If after every storm you warn against there are only three drops of rain, I will eventually not pay attention to it anymore. Therefore, it might be better to raise the threshold or to show graduated warnings." (E27 #00:07:24).*

More positively, one participant highlighted the desired learning effect of the plugin: *"I can imagine that it is very good to learn what you have to pay attention to. At some point when you have enough practice, you have taken on the same policies." (E14 #00:06:14).* Additionally, the participants had some interesting ideas to improve the plugin. For instance, it would be a helpful feature to display a link to the scientific sources of the chosen indicators. This is in the spirit of our white-box approach and would enhance transparency and objectivity. Furthermore, as the plugin does not include checks on videos and images, that should be pointed out to the user.

Perceived Autonomy. When we asked the participants to what extent they felt autonomous or patronized when assessing the tweets, all 16 participants regarded not to feel patronized at all. They highlighted the neutrally phrased additional information and the fact, that the plugin does not decide if a tweet contains fake news: *"I did not find it patronizing at all, especially because the explanations are written very neutrally. (…) You still must keep thinking for yourself. The plugin says there might be an indicator, but it does not have to be fake." (E14 #00:06:58).* The perceived autonomy was also enhanced by the configuration feature: *"When punctuation is not a criterion for me, I can just switch it off" (E27 #00:09:07).*

Plugin Features: Usability and Layout. All but one participant managed to interact with the additional information pop-up intuitively and very fast. Five participants noted explicitly that the additional information was helpful and necessary to understand the indicators completely. The configuration-feature has proven to be a substantially important aspect to enhance the perceived autonomy already in the two first studies using low-fidelity prototypes. The fully implemented version was used effectively and

intuitively by all 16 participants of the third study: *"Wow, that is easy! I do not have to think at all" (E21 #00:02:25).*
12 out of 16 participants managed to toggle the textbox containing the detected indicators straightaway. On the other hand, four participants struggled initially to find the correct button and tried a Twitter-internal button before toggling the textbox successfully. Most participants valued the feature since it can make the feedback more compact. The highlighting-feature was appreciated a lot by all participants. Some participants even said it was the most helpful component of the plugin as it helps the user to comprehend all warnings. Hence, the feature is central to our white-box approach. Using the highlighting-feature, all participants were able to match a specific warning to the correct referring part of the tweet successfully. All but two participants stated to like the layout of the plugin. While eleven out of 16 participants said it was noticeable enough, four were undecided and one found it was too noticeable. To avoid misunderstandings, participants suggested to add a mouseover-effect to the icon which appears when no implemented indicator was detected in a specific tweet.

5.3 Concluding Design Implications

Considering the presented results regarding perceived helpfulness, autonomy and usability of the plugin, we present five design implications to enhance the value of an indicator-based white-box approach to support users on Twitter in dealing with fake news. Those implications were extracted from statements that were mentioned particularly often or highlighted as very crucial by the participants. They support the view of existing studies, for instance by highlighting the relevance of transparent information (e.g. [6]) and the minimization of false alarms [27]. Moreover, they expand existing knowledge by a first scientific contribution on how to successfully develop an indicator-based white-box approach in that specific context.
Personalization to enhance autonomy. The configuration-feature is substantially important to enhance the perceived autonomy of the users and to prevent reactance. Our test subjects endorsed the possibility of deciding for themselves, on which indicators the tweets should be checked.
Assisting with transparent and objective information. The indicators need a detailed description that explains why they are relevant with regards to fake news. According to our test subjects, it is crucial, that the descriptions are formulated in a politically neutral and objective way. Adding a link to the associated scientific study can increase credibility in the spirit of our white-box approach. Warnings should always clarify that are assisting but do not replace the users' own assessment. Furthermore, the users must be aware of what functionalities the plugin does not include (e.g. our plugin does not examine images and videos).
Unambiguousness of warnings. Highlighting parts of the tweet in a mouseover-effect concerning a displayed warning was rated as one of the most helpful features of the plugin. Matching specific warnings to the correct referred part of the tweet is central to enhance the desired learning effect.
Personalized noticeability. The toggle-feature of the warnings was rated positively by the test subjects. Since the icon still gives a simple visible feedback, it is a pleasant way

of making the plugin more compact and adjusting it to the users' preferences concerning noticeability.

Minimizing false alarms. Minimizing false positives is crucial to prevent users from not paying attention anymore or uninstalling the plugin before having achieved a learning effect. Therefore, the threshold of warnings should not be too low. Some participants see advantages in showing graduated warnings in different colors.

6 Discussion & Conclusion

Dealing with fake news has proven to be one of the big current challenges in society and politics. Studies have shown that there is a need for assisting tools to support users on social media. There has been previous research on using machine learning algorithms to detect and label fake news. For example, Gupta et al. [12] present a browser-plugin to automatically assess the credibility of contents on Twitter. Further approaches (e.g. Fake News AI) use machine learning techniques as well. Other approaches are based on whitelists or blacklists (e.g. B.S. Detector) to detect fake news. The usage of black-box approaches though is not able to give reasons for its decisions and therefore, it runs the risk of creating reactance. In our eyes and according to other studies ([6], [7]), improving media literacy is a crucial strategy to help users dealing with fake news. Therefore, white-box approaches are necessary. However, all presented plugins, applications and approaches are based on black-box methods. Although the smartphone application Fake News Check can give transparent reasons on why contents might be fake, it does not automatically check for indicators and it comes with a big effort for the user.

Our scientific contribution is to theoretically explore the potential of an indicator-based white-box approach to assist users on Twitter and more practically to design, implement and evaluate a consistent browser-plugin as an artifact regarding to the design science approach. The plugin includes a warning with regards to six easily comprehensible and politically neutral indicators for fake news, further information about every indicator and a configuration-feature to support personalization. To answer our first research question (*How can we provide a transparent, politically neutral and objective assisting tool for users of social media?*), the empirical findings in Section 5.3 reveal that our indicator-based white-box approach to support users on Twitter in dealing with fake news can be considered suitable, applying the following five design implications: personalization to enhance autonomy, transparent and objective information, unambiguousness of warnings, personalized noticeability and minimization of false alarms. Moreover, we intended to answer the second research question: *Does a white-box approach counteract reactance and encourage a learning effect?* Our study shows that our white-box approach is a promising way to support users on social media without creating reactance but encouraging a learning effect and can therefore be considered a suitable alternative to black-box approaches.

Following our concept of design science, we intend to evaluate the newly suggested features (e.g. graduated warnings) in the future. Moreover, we intend to integrate further relevant user groups in our evaluation. In addition to our qualitative studies, a quantitative study is desirable to guarantee an evaluation on a larger scale. On the other hand, it would be interesting to examine if there is a beneficial way to combine our white-box tool with the features of a machine learning approach to receive the advantages of both methods, namely the transparent and easily comprehensible indicators of our approach and the accurate classifications of black-box approaches.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. Journal of Economic Perspectives 31, 211–236 (2017)
2. Reuter, C., Hartwig, K., Kirchner, J., Schlegel, N.: Fake News Perception in Germany: A Representative Study of People's Attitudes and Approaches to Counteract Disinformation. 14th International Conference on Wirtschaftsinformatik (WI 2019) (forthcoming)
3. Forbes: Can 'Fake News' Impact the Stock Market?, https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market (Accessed: 5.05.2018)
4. Reuter, C.: Information Technology for Peace and Security - IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace. Springer Verlag, Wiesbaden (Germany) (2019)
5. Asadullah, A., Kankanhalli, A., Faik, I.: Understanding Users' Intention to Verify Content on Social Media Platforms. 22nd Pacific Asia Conference on Information Systems, 251 (2018)
6. Müller, P., Denner, N.: Was tun gegen „Fake News"? (in German). Report. Friedrich-Naumann-Stiftung für die Freiheit. (2017)
7. Stanoevska-Slabeva, K.: Teaching Social Media Literacy with Storytelling and Social Media Curation. 23rd Americas Conference on Information Systems (2017)
8. Aufderheide, P.: Media Literacy. A Report of the National Leadership Conference on Media Literacy. ERIC (1993)
9. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. Management Information Systems Quarterly 28, 725–730 (2008)
10. Viviani, M., Pasi, G.: Credibility in social media: opinions, news, and health information—a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7, e1209 (2017)
11. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. In: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, p. 82 (2015)
12. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: TweetCred: Real-Time Credibility Assessment of Content on Twitter. 6th International Conference on Social Informatics, 228–243 (2014)
13. Rehm, G.: An Infrastructure for Empowering Internet Users to handle Fake News and other Online Media Phenomena. In: International Conference of the German Society for Computational Linguistics and Language Technology, pp. 216–231 (2017)
14. Pariser, E.: The filter bubble: How the new personalized web is changing what we read and how we think. Penguin (2011)
15. Kim, A., Dennis, A.R.: Says Who?: How News Presentation Format Influences Perceived Believability and the Engagement Level of Social Media Users. 51st Hawaii International Conference on System Sciences, 497 (2017)

16. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology 2, 175 (1998)
17. Kahne, J., Bowyer, B.: Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. American Educational Research Journal 54, 3–34 (2017)
18. Mihailidis, P., Viotty, S.: Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society. American Behavioral Scientist 61, 441–454 (2017)
19. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web, pp. 675–684 (2011)
20. Stieglitz, S., Bruns, A., Krüger, N.: Enterprise-related crisis communication on Twitter. 12th International Conference on Wirtschaftsinformatik (WI 2015), 917–932 (2015)
21. Baumann, A., Krasnova, H., Veltri, N.F., Ye, Y.: Men, Women, Microblogging: Where Do We Stand? 12th International Conference on Wirtschaftsinformatik (WI 2015), pp. 857–871 (2015)
22. Morris, M.R., Counts, S., Roseway, A., Hoff, A., Schwarz, J.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proceedings of the ACM 2012 conference on computer supported cooperative work, pp. 441–450 (2012)
23. Wanas, N., El-Saban, M., Ashour, H., Ammar, W.: Automatic scoring of online discussion posts. In: Proceedings of the 2nd ACM workshop on Information credibility on the web, pp. 19–26 (2008)
24. Weimer, M., Gurevych, I., Mühlhäuser, M.: Automatically assessing the post quality in online discussions on software. In: Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, pp. 125–128 (2007)
25. Weerkamp, W., Rijke, M.: Credibility improves topical blog post retrieval. Proceedings of ACL-08: HLT, 923–931 (2008)
26. van Someren, M.W., Barnard, Y.F., Sandberg, J.A.C.: The think aloud method: a practical approach to modelling cognitive. Academic Press, London (1994)
27. Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L.F.: Crying Wolf: An Empirical Study of SSL Warning Effectiveness. 18th USENIX security symposium, pp. 399–416 (2009)

Application of Process Mining Techniques to Support Maintenance-Related Objectives

Richard Horn[1], Patrick Zschech[1]

[1] Technische Universität Dresden, Business Intelligence Research, Dresden, Germany
{richard.horn,patrick.zschech}@tu-dresden.de

Abstract. The variety of data types generated in manufacturing environments leads to a situation where data-driven approaches for analytical maintenance support no longer have to be limited to the equipment level, but rather can be extended to further perspectives. To this end, this paper examines how process mining (PM) as an approach to extract knowledge about process-related relationships can be applied to support maintenance-related objectives. Our research is carried out by using exemplary data from a manufacturing company, where we successively take different data attributes from various source systems into account and apply selected PM techniques to demonstrate their applicability. As a result, we showcase how different insights can be provided, such as the analysis of a machine's internal behavior, examination of error dependencies across multiple production steps, determination of a machine's relevance within the equipment network or the discovery of bottlenecks regarding frequencies, cycle times and costs.

Keywords: Data-Driven Maintenance, Maintenance Objectives, Process Mining, Data Mining, Business Analytics

1 Introduction

In no other sector, more data are generated than in the field of manufacturing, ranging from process control and production status records to condition monitoring data of the overall equipment [1]. The variety of machine-generated data provides a vital asset for industrial maintenance, where approaches like condition-based maintenance (CBM) are applied for diagnostic and prognostic purposes to guarantee high reliability, low environmental risks and human safety [2-4]. Such approaches primarily focus on the equipment level and make use of data mining (DM) algorithms like clustering and classification techniques [5]. However, due to the diversity of manufacturing data, there are also other methodical approaches which allow a consideration from more distinct perspectives. Such further perspectives could be provided, for example, by applying techniques from the field of process mining (PM), where sequentially generated event data are utilized to extract hidden knowledge about process-related relationships and patterns [6]. In this way, it possible to obtain insights that are no longer based only on an instance level, but rather extend to an inter-unit consideration, making this approach also interesting for maintenance questions beyond a scope on the equipment level.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

While PM has already been applied successfully in various domains like transportation, healthcare, banking, retail or education [7], an application to the field of industrial maintenance is rather scarce within the current literature. For this reason, our paper aims to demonstrate its applicability within this particular context and addresses the following research question (RQ):

RQ: *How can process mining techniques be applied for knowledge extraction in manufacturing environments to support maintenance-related objectives?*

To carry out our research, we used exemplary data records from a manufacturing company, where we successively took data attributes from various source systems into account and applied PM techniques derived from the literature. Following this approach, the remainder of this paper is structured as follows: In Section 2, we provide the conceptual background for both disciplines of interest. We then showcase the applicability of PM techniques to support maintenance-related objectives in Section 3. Finally, we draw a conclusion and give an outlook for further research in Section 4.

2 Conceptual Background

2.1 Industrial Maintenance

Industrial maintenance can be understood as a broad field with many accentuations and facets. Therefore, a consideration of various maintenance definitions is required to derive central objectives for which analytical approaches based on different data assets can provide a methodical contribution.

In general, maintenance can be defined as a combination of all administrative and technical activities that are required for preserving the desired operating condition of the production equipment [8]. This statement can be extended by the British Standard Institution including the activities to restore the operating condition of production equipment [9]. Likewise, the guarantee on plant availability as well as part of the plant safety through resilient systems are connected to those objectives [10, 11]. Additionally, the DIN EN 13306 includes the aspects cost efficiency, environmentalism and product quality [12]. [13] augments this list with the efficient use of resources, which have divergent manifestations. Thus, [14] distinguishes “main resources” and “other resources”. Conclusively, the Maintenance Engineering Society of Australia announces “(…) that maintenance is about achieving the required asset capabilities within an economic or business context” [15] and specifies the optimization of production equipment as another aspect of maintenance. As a result, nine different core objectives of maintenance can be extracted as listed in Table 1, which will be referenced in the following by referring to their respective objective identifier (OID).

The actual maintenance execution is then carried out via different programs like time-based or data-driven concepts [2, 4], whereas our focus is on a data-driven support. Given the strong emphasis on the equipment level, the majority of data-driven concepts like CBM or predictive maintenance primarily concentrates on a particular unit of interest and thus employs classical DM techniques from statistics and machine learning [5]. This includes, for example, cluster analyses to detect unusual machine behavior, classifiers to determine heterogenous fault modes or artificial neural networks to predict a machine's remaining useful life [4, 5, 16]. However, due to the diversity of generated data in manufacturing environments, it is also possible to establish further analytical perspectives. As specified by our research goal, this will be demonstrated by applying techniques from the field of PM.

Table 1. Core Objectives of Industrial Maintenance

OID	*Objectives*	*References*
O1	Increase of machine lifetime	[8-11]
O2	Optimization of production equipment	[15]
O3	Retention or increase of product quality	[12]
O4	Minimization of machine downtimes	[8-11]
O5	Guarantee of safety	[10]
O6	Reduction of risk of failure and damage	[8-11]
O7	Efficient use of resources	[10, 13, 14]
O8	Retention of environment protection	[10, 12]
O9	Increase of cost efficiency	[12-15]

2.2 Process Mining

PM is characterized as a young research area that has been researched in the information systems domain in the last decade and established as a connection between business process management on the one hand and data science/business analytics on the other hand [6]. In general, PM tries to gain, aggregate and visualize both company- and process-relevant information by evaluating and analyzing great amounts of data in the sense of so-called event logs. Up to now, PM has been used in diverse fields like retail, education or healthcare, thus showing wide application areas. For example, while [17] and [18] use PM for the maintenance of web pages and their optimization and improvement, [19] tries to achieve personalized learning based on students' performance data. Furthermore, [20] show how PM can be used to identify deviations in healthcare processes from existing policies and best-practices, whereas [21] show a possibility to use PM to discover the customer fulfillment process in a telecommunication company.

Overall, there are three components of PM, each specialized on one remit. First, *process discovery* handles the creation of process models that are representative in behavior towards the underlying event logs. Hence, process discovery can be seen as the most relevant but also most difficult task of PM and offers many possibilities because of divergent process perspectives, such as the organizational or time-based perspective. By contrast, *process conformance checking* focuses the conformity of process models with the operative behavior observed within the event logs [6]. The aim of this comparison is to detect similarities, differences and deviations to evaluate "if actual processes follow prescribed behaviors or rules" [22]. As a last type of PM, the extension of existing event logs by additional attributes is characterized as *process enhancement*. These attributes define new perspectives and analytical possibilities in and on the process, whereby PM is opened for potential adjustments to reach subjective goals [6].

3 Maintenance Support with Process Mining

The following section demonstrates the applicability of PM towards the support of maintenance-related objectives. For this purpose, exemplary anonymized data from a manufacturing company producing car tires were used. The production process extends over various steps that are carried out on distributed machines. For the application of PM techniques, data from different source systems had to be brought into a standardized event log structure [6]. Here, we started with the minimum of required attributes at the lowest level and then successively considered additional attributes and levels to apply further perspectives. Relevant data were extracted from systems at different application layers, such as energy meters, programmable logic controllers, a manufacturing execution system (MES) or an enterprise resource planning system (ERP). For demonstration purposes, data samples from all systems were filtered and strongly simplified. Thus, the focus was primarily set on the overall feasibility instead of quantitively evaluating specific scenarios with regard to the current status and possible improvements.

3.1 Starting with the Minimum of Required Data

The minimum required data for PM is defined by the existence of two attributes: i) an *activity* that refers to an event class of interest and ii) a *case ID* that relates each event to a particular process instance. Since most PM tools also require a *timestamp* to ensure the event order, this attribute was integrated as well. With this basic data, it was possible to perform *process discovery* to investigate procedural structures at different levels [6].

At the lowest level, process activities can be defined by events derived from different machine components like logical elements, transistors or switching circuits. This allows to track consecutive states of a single machine and derive insights into a system's internal behavior without considering the entire production process. Thus, irritants like a high number of skipped events in machines or increased operation time can be detected at an early stage to trigger proactive maintenance actions and prevent further damage.

At a next level and this will remain the primary focus in the following considerations, process activities can be defined at the scope of the actual manufacturing process with all its steps and sub-steps mainly derived from MES data. This leads to multiple advantages as it becomes feasible to analyze the entire manufacturing setting regarding frequencies and durations of steps/sub-steps as well as variants, whereby one variant describes a specific path through the production process [23]. As a result, bottlenecks, time per step or variant and many other process statistics can be evaluated.

In an analogous manner, it is also possible to examine event data generated by the execution of maintenance tasks. This allows to reconstruct and critically analyze the procedures of operational maintenance activities. However, since most maintenance tasks are still executed manually without leaving extensive digital traces - as in the particular case where only scarce information could be extracted from the ERP - the focus will remain on the investigation of existing machine and production data.

Subsequent to the discovery of process structures, the results can be used to realize the *process conformance checking,* following [6]. As already stated, this allows to explore deviations and commonalities between the intended/documented processes on the one hand and the real as-is processes on the other hand. Moreover, it is possible to determine if specific process steps exceed predefined control limits like excursion rates, cycle times or a predefined amount of process instances to run on one single machine.

Overall, in terms of the maintenance objectives regarding machine lifetime (O1) and equipment optimization (O2), all those insights can be used to reveal unintended behavior of the production process at different abstraction levels and thus help to schedule maintenance activities proactively to prevent machine failures. Furthermore, this methodical approach offers the possibility to validate performed maintenance actions and their impact by comparing the process performance before and after an intervention.

3.2 Including Machine Attributes to Enrich Process Activities

While in the preceding scenario the activities of the production process were primarily considered in isolation, the database can further be enriched with resource information in the sense of *process enhancement*. Therefore, the process steps to be explored are subsequently not only specified by the production activities, but also by the machines performing those activities. Thus, after the *process discovery*, it can now be observed that all process steps consist of activity/machine-compositions, as shown in Figure 1, where the first step *"PSc3e047c7"* is performed by the machine *"S_636gt"*.

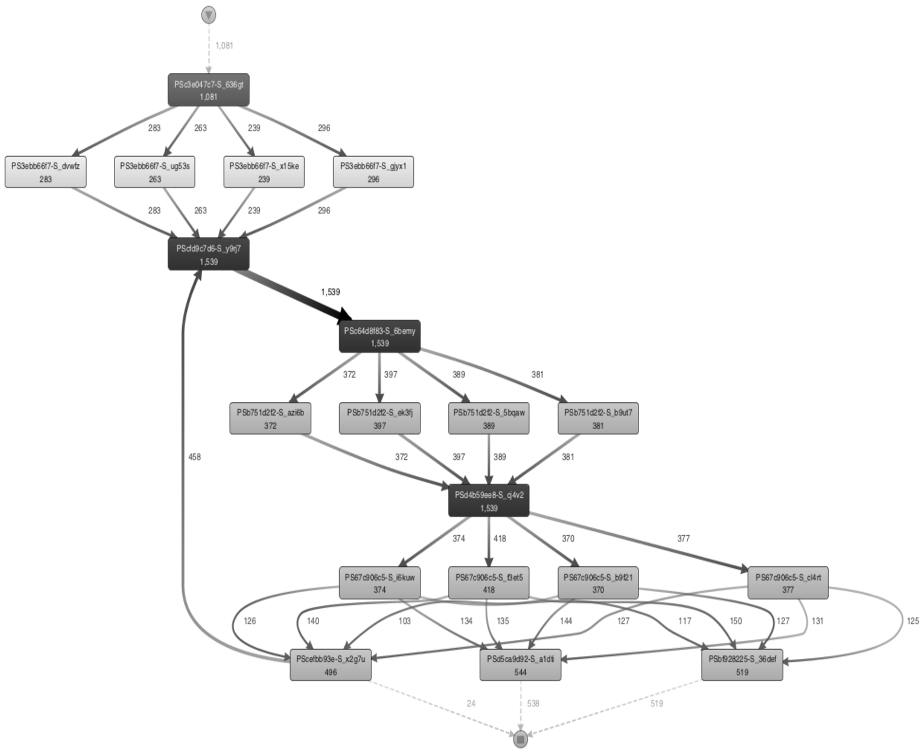


Figure 1: Simplified Production Process with Machine Data (Recorded in Disco)

With this representation, the workload of the machines can be expressed by frequency numbers and different color schemes indicating the different intensities. Thus, it is possible to detect machines with high workloads that are possibly more exposed to continuous stress. Those machines often occupy central positions within the production flow (like in the first, third, fourth and sixth step in Figure 1) and therefore require more attention to proactive maintenance, since a failure may disrupt the entire process (O4).

Similarly, the distribution of the workload between different machines performing the same production step can be examined. For example, the second step is executed by four different machines in parallel. However, while machine *"S_gjyx1"* was running 296 instances, unit *"S_x15ke"* was only executing 239 instances. In combination with scheduling information, these insights can be used for an optimal allocation of process instances to machines, with the goal to equalize the number of processed products (O2) and thus to lengthen the machines' lifetime (O1).

Moreover, by not only focusing on a single activity level but also considering their transitions, different process variants can be evaluated in terms of distinct production paths. Here, the activity/machine-composition offers the advantage to gather insights about sequential machine interdependencies. As such, it can be assessed, for example, whether some incidents primarily occur within process paths where certain machines were previously involved. This can also be relevant for quality assurance (O3), where specific machine combinations throughout the production flow may lead to different quality levels of the final product (c.f. Section 3.5).

Another perspective can be provided by analyzing durations of activities and transitions using timestamps instead of solely concentrating on frequencies. Analogously, time attributes can directly be incorporated in the process model to detect activity/machine-combinations with high cycle times in terms of bottlenecks and inefficiencies (O2). Simultaneously, exceptionally long cycle times are an indicator for faulty machine behavior to trigger further inspections and respective interventions.

3.3 Implementing and Expanding the XES-Lifecycle Extension

The exploration of failure events within the process is another relevant facet of maintenance support. For such a consideration, the given event log requires an additional attribute to document the production status of each step.

This can be implemented by employing the XES-Lifecycle Extension as introduced by [24], which provides predefined values like *"start"*, *"complete"*, *"suspend"*, *"resume"* or *"pi_abort"*, with the last value defining an interruption of the current process instance. It is conspicuous though, that with these values only the possibility is given to analyze interrupting machine errors and suspending errors. However, with regard to errors that do not interrupt the production process, this framework requires an extension. Thus, we created an additional state, called "*fail_complete*", to describe the completion of a process step while errors take place in execution. This allows to focus on failure events in the process analysis and generates high profit for subsequent analysis, such as tracking overall failure times. For example, the PM tool Celonis allows to define domain-specific key performance indicators (KPI) that can be monitored live during process execution. By using this feature in combination with the newly created state, KPIs for the number and the ratio of failure events in a single production equipment can be defined. This live-monitoring of the production process combined with the historical process data gives some indication of past and future machine behavior. For instance, this enables to determine the average number of internal errors until the process instance is finally interrupted by a serious error (O4).

Moreover, also the duration of failure events is of high relevance for maintenance. At this point, there are multiple ways to achieve the documentation of these downtimes. Enhancing the event log by inclusion of an additional attribute containing these data represents one way to accomplish this goal. However, this approach contradicts with one of the fundamental principles of PM, as the downtimes exhibit a direct reference to the processing machine and not to the activity. Another option arises when saving the desired information in an additional event log defining the production equipment as events. But this leads to a loss of the process perspective. Therefore, it's necessary to calculate the downtimes by using existing data and to add another dimension to the predefined time perspective. Based on the assumption that the downtime is calculated by the timestamps of the failure and the reactivation of the machine, the dead times can be computed by using the pseudocode depicted in Figure 2. In contrast to the previous approaches, this allows to view dead times directly in the process and to analyze them in case specific or cumulative way by using all advantages of *process discovery*.

```
# get_breaktime:
n = max(#CaseID)
FOR i=0 TO n-1 DO
  current_machine = #Machine(e)[i]
  current_LC_transition = #LC(e)[i]
  IF current_LC_transition == "stop" THEN
      FOR j = i+1 TO n  - 1 DO
          next_machine = #Machine(e)[j]
          IF current_machine == next_machine THEN
              #Starttimestamp(e)j - #Endtimestamp(e)i
          END IF
      END FOR
  ELSE
      continue
END FOR
```

Figure 2: Pseudocode for Calculating Downtimes

Apart from this, dependencies between different errors in the process are focused. For the discovery of such dependencies, different techniques from statistics and machine learning can be used, such as support vector machines or logistic regression [25]. Exemplarily, we use the logistic regression that allows to receive direct failure probabilities. Transferred to the present context, this enables to determine if the probability for a specific machine documenting an error increases, when another machine earlier in the process also documented an error. It could be found, for example, if station *"S_636gt"* documents an error, then station *"S_dvwfz"* is 2.56 times more likely ($B = 0.94, p < 0.001$) to also write an error in reference to the state in which station *"S_636gt"* would not have documented one ($\mathrm{X}^2(1) = 23.84, p < 0.001$). In this way, failure dependencies in the process can be revealed and provided for decision support by directly visualizing this information within the process model.

This particular scenario illustrates the successful combination of PM and DM techniques. PM first aims at uncovering sequential patterns to identify process complexities and distinct paths. This allows the detection of discrepancies between the desired and actual process behavior, for example, in the form of anomalous paths, bottlenecks or deviating process performance. Once such discrepancies are detected, DM techniques can then be used to identify non-local and multi-causal effects that possibly span over multiple process steps [26]. Referring to machine failures during the execution of production processes, it is possible to determine patterns which classify a specific sequence of error events. Thus, the option arises to predict the breakdown of a machine based on the error events that occurred in previous stages, which allows to interfere in the process at the right time and to perform tasks to prevent upcoming machine failures (O1, O4).

3.4 Adding the Organizational Extension

Another perspective provided by PM is the organizational view, with the goal to extract and visualize social network structures between different entities involved in a process environment [7]. In this context, a social network consist of i) nodes representing organizational units and ii) relationships representing the connections between those units [6]. Transferring this approach to the current setting, the first part can be defined by the production equipment, whereas the second part is characterized by ingoing- and outgoing connections according to the manufacturing process. To achieve this kind of networks, the XES Standard defines the "organizational extension", encompassing the name or the ID of the respective operator in the event log [24].

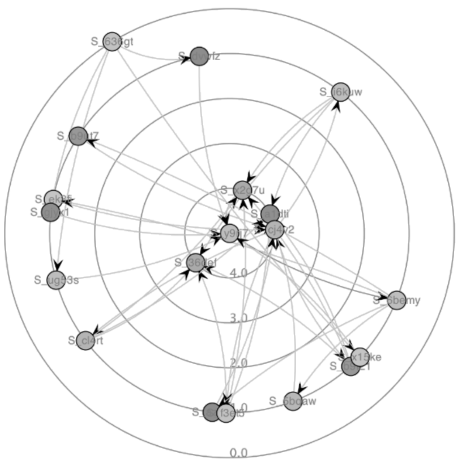


Figure 3: Network of Production Equipment (Recorded in ProM)

Besides the creation of a visual network (c.f. Figure 3), the machines can additionally be assessed according to their importance within the network using centrality measures. Following this, the most valuable machine in the network is characterized by the highest number of connectors [27]. Focusing on the reduction of risk and failure damage (O6), it can be concluded that theoretically station *"S_y9tj7"*, positioned in the center, is of high relevance for the production process, as its failure may influence different other machines which could lead to a high capacity loss. Nevertheless, this approach only indicates a high equipment relevance based on relations to other production equipment, whereas ramifications of failures, like emission of harmful substances, are not included in this risk estimation.

3.5 Integration of Domain-Specific Attributes

When executing production steps, machines usually consume different input resources like energy and materials while simultaneously producing waste and other by-products. As such, the event log can further be enriched with attributes in terms of measured values and specific KPIs to illustrate the resource flow at a machine and process level.

Data attributes can be gathered from various systems, as in the current case where power consumption was measured via energy meters and material flows were collected from the MES. However, event characteristics are generally not limited to those aspects, but rather can extend to any other subject of interest, such as further environmental indicators [28] or quality-related attributes [29].

The integration of attributes can be done in two ways: On the one hand, it is possible to simply add them to the event log by appending it as a property to the corresponding events. On the other hand, an explication of an additional extension to the XES-Standard is possible to achieve a certain level of standardization by specifying concrete definitions of the attributes in the extension definition. This also permits to create domain-specific extensions for a complete set of properties, which then can be in the center of further analysis with a predefined prefix for their identification [24].

Once the event log is enriched with further attributes, they can be examined at the different levels addressed previously, i.e. for i) the overall production process, ii) individual production steps, iii) individual machines, iv) activity/machine-combinations, or v) distinct paths. In this way, it is possible, for example, to monitor resource efficiencies (O7), but also to track material flows in terms of waste and harmful materials (O5, O8).

In addition, it is also advisable to not only consider each attribute in a univariate manner, but rather to analyze them in combination with each other along the entire manufacturing process. Here, the combination with DM techniques proves to be helpful again, where each attribute either serves as a target variable of interest or as an input feature to predict the respective output [26]. This makes it possible, for example, to determine the relationship between the final product quality, measured at the end of a process, and the machine and process attributes at each individual production step (O3).

3.6 Evaluating the Cost Efficiency

In a production process, there are many different types of costs that need to be regarded when evaluating the overall cost efficiency. Therefore, an individual cost model needs to be applied to determine which parameters to include to the event log. In the given context, the costs for one process activity e in the set of all process activities E will be calculated with $\#_{TC}(e) = \#_{EC}(e) * \#_{t}(e) * EP + \#_{MC}(e)$ with TC as total costs of event $e \in E$, EC and EP as energy consumption and its current price, t as duration and MC as material consumption. When enriching the event log with the not yet included (external) data, it is noticeable that, due to the event centric focus of PM, there is a loss of information or costs regarding the event transitions of the process model. These transitions also need to be regarded as they are in the time or frequency perspective of a process. It is conspicuous that there are costs like the energy consumption in standby mode of a machine, required energy to get in production ready state or the consumption for cooling down in the time between two process steps in a production process. In conclusion, there are costs right before and after executing a specific task that are not yet regarded by an event centric focus. By further adding the cost calculation at the start and end timestamp of a process activity, possibilities arise to determine these unregistered costs and to provide a cost perspective of the process. Therefore, the costs before activity execution can be determined by subtracting the costs at the start of an event $\#_{StartCosts}(e_i)$ from the costs at the end of the previous executed event $\#_{EndCosts}(e_{i-1})$, whereas the costs after execution are determined by the subtraction of all cost attributes of an event: $cae(e_i) = \#_{EndCosts}(e_i) - \#_{ActivityCosts}(e_i) - \#_{StartCosts}(e_i)$. Given these attributes, it can be stated that the overall costs of a transition ct between events can be calculated by adding the pre-execution costs of an activity to the post-execution costs of the activity before that:

$$ct(e_i, e_{i+1}) = cae(e_i) + \left(\#_{StartCosts}(e_{i+1}) - \#_{EndCosts}(e_i)\right) \mid e_i \succ e_{i+1} \qquad (1)$$

In conclusion, a cost perspective on the process can be realized that allows to gather all information available through *process discovery* and to achieve a long-term validation and possible predictions of future production costs. Therefore, cost referred bottlenecks and cost intensive areas of the process can be identified (O9). Moreover, by focusing on process variants, optimization potentials can be realized, e.g. by distributing process instances to less cost-intensive machines (O2).

4 Conclusion and Outlook

The goal of this paper was to demonstrate the application of PM techniques in manufacturing environments to support maintenance-related objectives. Using successively more data attributes from various source systems at different application levels, it could be shown how insights from multiple perspectives were achievable. This includes, for example, i) the discovery of a machine's internal behavior by tracking its consecutive states, ii) the identification of error dependencies across multiple production steps by combining PM with DM techniques, iii) the determination of a machine's relevance within a network by using organizational mining, iv) the examination of bottlenecks by analyzing different process levels (e.g., process paths vs. single activities) with regards to time, frequency and cost indicators, or v) the evaluation of a machine's input and output efficiencies based on domain-specific attributes. Consequently, all maintenance objectives (ranging from aspects like lifetime extension and equipment optimization to resource efficiency and quality assurance) could be addressed to some extent.

While the focus at this stage was mainly on the demonstration of the overall applicability and feasibility, future work will be devoted to a more detailed examination in the sense of a quantitative evaluation. As such, it is planned - in accordance with given confidentiality restrictions - to provide more details about the data samples in each analysis scenario and then quantitatively discuss the results of applied approaches. For example, such an in-depth consideration could be based on a large number of summary statistics and indicators on the recorded behavior of machines and processes, which are then compared with the expected results from technical experts to evaluate how the generated insights can be used for better decision support.

Likewise, further work will place greater emphasis on the necessary data preparation steps, as this was largely ignored in the current study. For demonstration purposes, different data samples were filtered and strongly simplified. However, the integration of different data attributes from heterogenous systems into a standardized PM event log structure has to be considered as a challenging task [6]. For example, most of the involved manufacturing systems provide required event data only as by-products, which then have to be selected and merged at an adequate abstraction level on which a process flow is to be analyzed. Similarly, specific event attributes like energy or material consumption are often not recorded by default for each individual machine at each single production step. The challenge is therefore to assign such attributes in the event log according to their cause, thus enabling more fine-grained analyses. Consequently, those and many other aspects still remain as data preparation challenges [30], that may considerably hinder a successful PM application. However, once these steps are taken, our study showed a possible direction for how PM can provide multi-perspective insights to support decision making in maintenance as well as manufacturing settings in general.

References

1. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big Data: The Next Frontier for Innovation, Competition, and Productivity (2011)
2. Zschech, P.: A Taxonomy of Recurring Data Analysis Problems in Maintenance Analytics. In: 26th European Conference on Information Systems, Portsmouth, UK (2018)
3. Bousdekis, A., Magoutas, B., Apostolou, D., Mentzas, G.: Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. J. Intell. Manuf. 29, 1303–1316 (2018)
4. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Sign. Proc. 20, 1483–1510 (2006)
5. Accorsi, R., Manzini, R., Pascarella, P., Patella, M., Sassi, S.: Data Mining and Machine Learning for Condition-based Maintenance. Procedia Manuf. 11, 1153–1161 (2017)
6. Van Der Aalst, W.: Process Mining: Data Science in Action. Springer, New York (2016)
7. Van Der Aalst, W.: Intra- and Inter-Organizational Process Mining: Discovering Processes within and between Organizations. In: IFIP Working Conference on the Practice of Enterprise Modeling, pp. 1–11. Springer, Berlin, Heidelberg (2011)
8. Muchiri, P., Pintelon, L., Gelders, L., Martin, H.: Development of maintenance function performance measurement framework and indicators. Int. J. Prod. Econ. 131, 295–302 (2011)

9. CIBSE Guide M.: Maintenance engineering and management: A guide for designers, maintainers, building owners and operators, and facilities managers. Chartered Institution of Building Services Engineers (2008)
10. European Committee for Standardization: Business Plan CEN/TC 319 - Maintenance Executive Summary (2013)
11. Visser, J.K., Pretorius, M.W.: The development of a performance measurement system for maintenance. S. Afr. J. Ind. Eng. 14, 83–97 (2003)
12. DIN EN 13306: Instandhaltung - Begriffe der Instandhaltung. Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin (2015)
13. Kelly, A.: Maintenance and its management. Conference Communication (1989)
14. Ashayeri, J.: Development of computer-aided maintenance resources planning (CAMRP): A case of multiple CNC machining centers. Robot. Comput. Integr. Manuf. 23, 614-623 (2007)
15. Murray, M., Fletcher, K., Kennedy, J., Kohler, P., Chambers, J., Ledwidge, T.: Capability Assurance: A Generic Model of Maintenance. In: 2nd International Conference of Maintenance Societies, pp. 1–5. Melbourne (1996)
16. Bink, R., Zschech, P.: Predictive Maintenance in der industriellen Praxis: Entwicklung eines Prognoseansatzes unter eingeschränkter Informationslage. HMD. 55, 552–565 (2018)
17. Mihai, I.: Web Mining in E-Commerce. Annals of the University of Oradea, Economic Science Series 18, 959–963 (2009)
18. Poggi, N., Muthusamy, V., Carrera, D., Khalaf, R.: Business Process Mining from E-Commerce Web Logs. In: Business Process Management, pp. 65–80. Springer, Berlin, Heidelberg (2013)
19. Škopljanac-Mačina, F., Blašković, B., Skočir, Z.: Using Formal Concept Analysis for Student Assessment. In: 56th International Symposium ELMAR, pp. 1–4. Zadar, Croatia (2014)
20. Rovani, M., Maggi, F.M., de Leoni, M., van der Aalst, W.M.P.: Declarative process mining in healthcare. Expert. Syst. Appl. 42, 9236–9251 (2015)
21. Mahendrawathi, E.R., Astuti, H.M., Nastiti, A.: Analysis of Customer Fulfilment with Process Mining: A Case Study in a Telecommunication Company. Procedia Comput. Sci. 72, 588–596 (2015)
22. Aleem, S., Capretz, L.F., Ahmed, F.: Business Process Mining Approaches: A Relative Comparison. arXiv preprint arXiv:1507.05654 (2015)
23. Stocker, T., Accorsi, R., Rother, T.: Computergestützte Prozessauditierung mit Process Mining. HMD. 50, 92–103 (2013)
24. Günther, C.W., Verbeek, E.: XES Standard Definition. Fluxicon Process Laboratories (2014)
25. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
26. Hilbert, A., Zschech, P.: Process Analytics. WISU 45, 942–948 (2016)
27. Zhang, G., Duan, H., Zhou, J.: Network stability, connectivity and innovation output. Technol. Forecast. Soc. Change 114, 339–349 (2017)
28. Weiß, D., Müller, R., Lössl, S.: Umweltkennzahlen in der Praxis: Ein Leitfaden zur Anwendung von Umweltkennzahlen in Umweltmanagementsystemen mit dem Schwerpunkt auf EMAS. Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Referat Öffentlichkeitsarbeit (2013)
29. DIN EN ISO 9000: Quality Management Systems - Fundamentals and Vocabulary. Deutsches Institut für Normung e.V., Beuth-Verlag, Berlin (2015)
30. Van Der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., et al.: Process mining manifesto. In: International Conference on Business Process Management, pp. 169–194. Springer, Berlin, Heidelberg (2011)

How Voice Can Change Customer Satisfaction: A Comparative Analysis between E-Commerce and Voice Commerce

Daniel Kraus[1], Victoria Reibenspiess[1], Andreas Eckhardt[1]

[1] German Graduate School of Management and Law, Heilbronn, Germany
{daniel}@lorenzenkraus.de
{victoria.reibenspiess, andreas eckhardt}@ggs.de

Abstract. Voice commerce is a newly evolving e-commerce channel where consumers communicate with dedicated systems on smart speakers or other devices using their voice, in order to find products. This paper comparatively investigates factors for customers' satisfaction in voice commerce and e-commerce. Being the first study to scientifically analyze customer satisfaction factors in voice commerce and compare them with e-commerce, we conducted a survey with 178 consumers and used structural equation modeling for statistical hypotheses testing. The results show, that consumers have higher expectations in convenience for voice commerce than they have for e-commerce. Transaction process efficiency significantly influences satisfaction in voice commerce, but not in e-commerce. This research provides implications for future research on voice commerce strategy and system design.

Keywords: Voice Commerce, E-Commerce, Conversational Agent, Recommender Systems, Customer Satisfaction

1 Introduction

Since their introduction in 2014, the use of intelligent virtual assistants based on smart speakers like Amazon Alexa, Apple HomePod, Microsoft Cortana and Google Home is increasing [1]. Moar [2] estimates that there are currently 450 million voice assistant devices in the US, expected to reach 870 million by 2020. These systems make it possible to conduct a "zero-click" purchase in business to consumer (B2C) commerce scenarios. Communicating with the assistant using only their voice, consumers can formulate search queries and confirm purchase actions without the need to use common visual or typing interfaces. Electronic Commerce (e-commerce) experts label this scenario "voice commerce" and expect it to be one of the most important innovations to shape the next years of e-commerce development (e.g., [3-4]). E-commerce describes commerce conducted over electronic media, such as the use of the internet to facilitate and process business transactions [5]. Voice commerce as a subset of e-commerce provides consumers with computerized voice technologies

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

(e.g., speech recognition, voice identification, and text-to-speech) to execute these business transactions [6]. These systems involve natural language processing (NLP), intent recognition, speech synthesis, recommender systems and artificial intelligence (AI) technologies (e.g., [3],[5]).

Despite a long-standing research interest in customer satisfaction and loyalty factors for e-commerce applications (e.g., [6-7]) as well as on e-commerce using conversational text-based interfaces (e.g., [10]), specific research on e-commerce in a human-to-AI voice-based scenario is, however, sparse. Research related to customer satisfaction factors in voice commerce is entirely missing from current literature, as well as research aiming at possible differences in customer satisfaction factors (CSF) between e-commerce and voice commerce. Similar to mobile commerce (m-commerce) in comparison to e-commerce, voice commerce is subject to special restrictions and presents different opportunities and value proposition to customers. Therefore, it is likely that satisfactory factors for voice commerce might differ from those of e-commerce both in existence and importance.

To support voice commerce software design and implementation, managers need to know which factors influence customer satisfaction. While many CSF for e-commerce applications are known, it is difficult to ascertain factors for voice commerce from current literature. Therefore, our research question is:

RQ: *How do the influencing factors for customer satisfaction differ in voice and e-commerce?*

To identify customer satisfaction factors for voice commerce, we first review research related literature. Based on this review, we develop our research models regarding customer satisfaction and its predictors, consisting of four comparative hypotheses (cf., [11]). Following this, we describe our research design and methodology to empirically validate our models for both e-commerce and voice commerce. Afterwards, we analyze the data gathered by a survey using structural equation modeling and present our findings. Finally, the paper discusses theoretical and practical implications for management as well as presents limitations and gives directions for future research opportunities.

2 Theoretical Background and Hypotheses Development

E-commerce describes commerce conducted over electronic media. For example, Kwon and Sadeh [5] define e-commerce as the use of the internet to facilitate, execute, and process business transactions. However, in science the term is mainly used for electronic commerce conducted via computers and laptops, as opposed to mobile devices (e.g., [5]), although these devices also use the internet. Researchers label the latter scenario mobile commerce or m-commerce [12], defined as a subset of all e-commerce transactions [5].

One subset of e-commerce is conversational commerce utilizing neuro-linguistic programming (NLP) (e.g., [13]). Such interfaces can be either text-messaging or voice recognition systems [14]. One form of conversational commerce are commercial chatbots (e.g., [15]). The actual interaction is text-based, in which both human and machine generate written text to convey information [16]. Some commercial chatbots can also display product images and other visual information [17]. Animated or embodied agents (sometimes also called avatars) are conversational systems that provide a visual representation of the virtual agent in addition to a text or speech interface [13, 18]. Luger and Sellen [3] use the term conversational agent for an emergent form of dialogue system that is becoming increasingly embedded in personal technologies and devices. Galanxhi and Nah [6] define voice commerce as e-commerce involving computerized voice technologies: speech recognition, voice identification, and text-to-speech. In our context, we define voice commerce as a subset of e-commerce providing consumers with computerized voice technologies to facilitate, execute, and process business transactions (e.g. [6]).

2.1 Recommendation Complexity

Conversational recommender systems, like voice commerce and chatbots, converse with users to learn their preferences and incorporate feedback from users (e.g., [19]). Liang et al. [20] found that recommendation accuracy of these systems is positively linked to customer satisfaction. Xiao and Benbasat [21] point out that recommender systems can decrease the information overload facing consumers, as well as the complexity of online searches. For e-commerce applications, Xiao and Benbasat [21] investigate the usage of recommendation agents and created a complex interactive model of recommendation effectiveness, were product type and complexity play significant roles.

Recent research provides limitations for voice commerce using only auditory interfaces. The cognitive cost-benefit framework [22] predicts that consumers search less as media richness decreases because of higher cognitive effort for searches in low media richness environments [23]. E-commerce, due to its higher media richness and visual/text efficiency, usually presents a larger evaluation set [24]. Research by Maity and Dass [23] shows a negative impact of an "overwhelming amount of information" in low media richness channels, like voice commerce, compared to high media richness channels such as e-commerce and physical stores. This can also be applied to the presentation of recommendations, which are usually presented in the form of result lists, similar to normal search results [25]. However, to avoid exceeding user's information capacity and to reduce the time spent by a consumer to listen to recommendations, the complexity of recommendations in voice commerce can be reduced intentionally. Recommendation complexity can be subdivided into quantity of product recommendations and complexity of a single product presentation (i.e., the length and level of detail of the product description). Possibly, consumers appreciate more options than just a single one. In essence, these considerations lead to the assumption that voice commerce only supports a lower customer decision complexity. Therefore, we hypothesize:

Hypothesis 1 (H1): *Recommendation complexity has a larger effect on customer satisfaction in e-commerce than in voice commerce.*

2.2 Recommendation Personalization

In most current e-commerce platforms, search engines use recommendation features and personalize results for the consumer and also enrich them with data from social media [26]. Personalized recommendations are known to increase customer satisfaction and conversion rates, and to lower the size of the evaluation set [26-27]. The use of personalized recommendation agents generally reduces the number of products for which users want to retrieve detailed information [29]. Users of digital assistants expect a highly personalized system, as Chopra and Chivukula [4] report for Indian consumers.

If voice commerce benefits from a lower recommendation complexity, it implies a greater need for highly accurate recommendations, of which personalization is a main factor. A buying decision becomes easier if the user herself has made that same decision before, or if the system can draw upon preferences it knows about the user. Product-wise, customers are less likely to purchase high-involvement goods like a television or a dishwasher via voice commerce because of informational complexity involved. In contrast, it is more likely that customers have a tendency to purchase low involvement goods, as indicated by Maity and Dass [23] that customers are likely to undertake simple decision-making tasks on channels that incorporate low levels of media richness. Additionally, customers have a tendency to buy goods they have bought before. Personalization in recommendations can also be based on inferred or mentioned preferences from previous user-machine dialogue [30] or even based on learned body measurements (e.g., for clothing). Therefore, we hypothesize:

Hypothesis 2 (H2): *Personalized recommendations have a larger effect on customer satisfaction in voice commerce than in e-commerce.*

2.3 Convenience

Convenience is one of the most prominent CSF in e-commerce (e.g., [30-31]). Choi et al. [33] define convenience as the degree to which a person believes that navigating or engaging in transactions through e- or m-commerce is free of effort. Further they subdivide convenience into ease of use, ease of access, ease of understanding, usefulness and functionality [33].

A study by Chai et al. [34] found that most users preferred a commercial chatbot interface over a classic search interface as they liked the idea that they can express their needs in their language without being restricted to menu choices and that the computer does all the work for them [34]. For voice-based interfaces, Luger and Sellen [3] report that the principle use-case for the CA (conversational agent) was “hands free”, which was tied strongly to the theme of time-saving and convenience. This fits well to the previously mentioned idea that audio interfaces facilitate multi-tasking [35]. The efficiency and easiness of speech input is a value proposition that also plays a role. According to Luger and Sellen (2016), customers feel it is often easier and more convenient to use speech input than to type, one reason being that speech was felt to be faster. In their comparative studies, Choi et al. [33] and Cao et al. [12] found customers scored convenience higher for m-commerce than for compared to e-commerce. Since voice commerce should rank lower than e-commerce in media richness, convenience should be of greater importance. We hypothesize that customers have higher convenience expectations of voice commerce than of e-commerce:

Hypothesis 3 (H3): *Convenience has a larger effect on customer satisfaction in voice commerce than in e-commerce.*

2.4 Transaction Process Efficiency

Transaction time is a known CSF in e-commerce research (e.g., [36-37]). For example, Devaraj et al. [36] found that subjectively, excess time spent in the transaction process decreases satisfaction in e-commerce, whether it is spend on communication, searching and choosing or payment. Choi et al. [33] define the e-commerce CSF of "transaction process" as a combination of efficiency, total transaction time, clearness of the process and response time for each step. Their results show that these performance indicators very significantly in influencing satisfaction of different types of e-commerce.

Chatbot users frequently mention a high performance expectation, with subcategories of fast, efficient, and reliable [38]. Users expressed that the use of chatbot systems should reduce interaction time and increase efficiency [34]. In this context, they define efficiency as the number of clicks and the amount of time required obtaining the relevant information. By investigating task-oriented spoken dialog systems, Walker et al. [39] also found that a significant satisfaction factor is user's perception of elapsed time. According to research on users of current generation conversational agents, timesaving was a key related motivation to use these systems [3]. Therefore, we hypothesize:

Hypothesis 4 (H4): *An efficient transaction process has a larger effect on customer satisfaction in voice commerce than in e-commerce.*

3 Research Methodology

We conducted focus groups with e-commerce experts of an e-commerce consulting company in order to verify the importance of the listed constructs above (i.e., Recommendation Complexity, Recommendation Personalization, Convenience, and Transaction Process Efficiency).

To empirically test the proposed hypotheses, a survey was executed with the help of the crowd-sourcing platform Amazon Mechanical Turk (MTurk). Our questionnaire[1] consists of 39 questions, of which 13 serve demographical, 26 are related to the identified CSF and reflected both research models on e-commerce and voice commerce. Respondents have to answer the items for both types of commerce. We derived the questions for convenience from a study by Choi et al. [33] (e.g., "Ordering products **on websites**/*with my voice* is easy.", whereas the bold phrase represents the e-commerce and the italic characterizes the voice commerce construct).

[1] A comprehensive table of the measurement items can be accessed here: https://www.dropbox.com/s/ahovk0cfgcgihg5/Appendix_Studentstrack.pdf?dl=0

The measures for the construct transaction process efficiency are also adapted from Choi et al. [33] (e.g., "When ordering products on **a website**/*with my voice*, the process should take as little steps as possible."). Recommendation personalization as a construct was used by Komiak and Benbasat [28]. For example, we took items as "When ordering products on **a website**/*with my voice*, I benefit from product recommendations based on what I ordered before". Further, we adapted items for recommendation complexity (e.g., "When ordering products on **a website**/*with my voice*, I benefit from very detailed product recommendations.") based on product complexity in recommender systems [21]. Customer satisfaction was measured with items, such as "I am generally satisfied when ordering products **on websites**/*with my voice*." adapted from Chang and Chen [40]. To eliminate wording inconsistencies or comprehension problems we ran an independent pre-test with some participants, who were then excluded from the main survey [41].

The survey was restricted to only those US residents who have been consumers of both voice commerce and e-commerce systems in the previous three months. Therefore, we filtered out inappropriate participants before we conducted the main study. In a total 178 people answered the survey completely. Out of these 178, 53.9% were women and 44.9% men (1.1% did not give any information). The age distribution was: 25-44 years (70.2%), 18-24 years (17.4%), 45-64 years (9.6%), and 65 years and older (2.8%).

4 Data Analysis

In order to analyze the proposed research model and to validate the proposed hypotheses, the model has been transferred into a structural equation model [42]. For this examination the software IBM AMOS 21.0 was used to determine path influences. The suggested ratio of sample size to number of free parameters of 10:1, in order to reach trustworthiness, is fulfilled [43, 44].

4.1 Measurement Models

To begin with further data analysis, we calculated Cronbach's alpha to assess the internal consistency and reliability of the sub-scales. In the first iteration, some items showed low item-total correlation. All values calculated exceeded the recommended minimum value of 0.6, which indicated that the constructs show a high level of reliability [45, 46] (see Table 1).

We carried out a principal component analysis to identify component fit. Furthermore, we applied main component analysis as extraction method and Varimax (as our employed factors are not correlated) as rotation method (Kaiser-Normalization, convergence after 6 iterations). The model with four components fits well with an average loading of 0.82 and no cross loadings above 0.43, also indicating convergent validity. By calculation, the four factors account for 73.6 % of the total variance. The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was 0.68 and 0.69, representing a relatively good factor fit by exceeding the threshold value of 0.5 [47]. Bartlett's test of sphericity was significant ($p < .001$), indicating that correlations between items were sufficiently large for performing a factor analysis (compare table 2).

Table 1. Evidence of reliability

Model	*Construct*	*Items*	*Cronbach's alpha*
eCom	Convenience (EC)	3	0.62
	Recommendation personalization (ERP)	2	0.84
	Recommendation complexity (ERC)	2	0.72
	Transaction process efficiency (ETP)	2	0.62
vCom	Convenience (VC)	3	0.80
	Recommendation personalization (VRP)	2	0.83
	Recommendation complexity (VRC)	2	0.80
	Transaction process efficiency (VTP)	2	0.68

Table 2. KMO and Bartlett tests

Model	*Test*	*Indicator*	*Value*
eCom	KMO	Measure of sample suitability	0.68
	Bartlett	Approximate chi-square	425.461
		df	36
		Significance	< 0.001
vCom	KMO	Measure of sample suitability	0.69
	Bartlett	Approximate chi-square	634.8
		df	36
		Significance	< 0.001

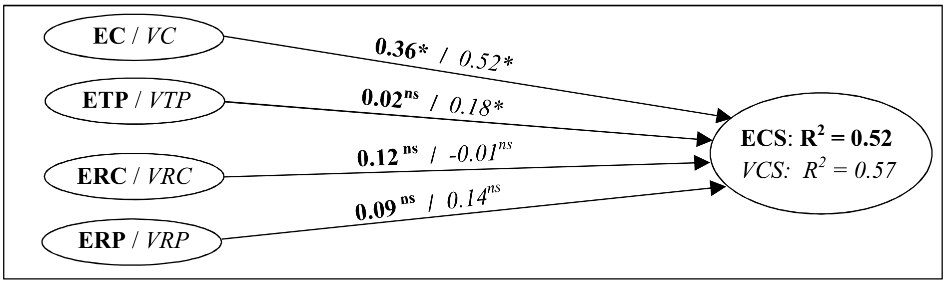
Tables 3 shows the factor correlation matrices for both models including composite reliability (CR) [48]. CR is above or near 0.7, except for transaction process efficiency. The latter achieved a CR value of 0.652 in the e-commerce model. The square root of the AVE is represented by the diagonal elements in table 3. The values show that the square root is bigger than each off-diagonal element [49]. We infer that there is an acceptable and logical extent of discriminant validity in the measurement model for all constructs.

Table 3. Factor correlation matrix

Model	*Construct*	*CR*	*ERP*	*EC*	*ERC*	*ETP*
eCom	ERP	0.835	**0.847**			
	EC	0.693	0.025	**0.659**		
	ERC	0.790	0.620	0.061	**0.816**	
	ETP	0.652	0.202	0.497	0.074	**0.699**
Model	*Construct*	*CR*	*VRP*	*VC*	*VRC*	*VTP*
vCom	VRP	0.838	**0.850**			
	VC	0.809	0.260	**0.766**		
	VRC	0.813	0.653	0.321	**0.829**	
	VTP	0.697	0.347	0.272	0.067	**0.735**

4.2 Structural Models

We created two structural models and performed an initial factor estimation using the maximum-likelihood method (see Figure 2). Straight arrows connecting each latent construct to customer satisfaction represent unidirectional effects, annotated by the standardized path coefficient. The total variance in customer satisfaction explained by the independent variables (R^2), which reflect the predictive power of the models, is above 50% in both models (0.52 in e-commerce and 0.57 in voice commerce).



^{ns}: not significant above $p < 0.05$ level, * : significant above $p < 0.05$ level; Bold = e-commerce model, Italic = voice commerce model

Figure 2. Structural model (e-commerce and voice commerce)

Table 4 shows the results for all predictors for ECS and VCS in the model. The effect of convenience is statistically significant in both e-commerce and voice commerce. Based on the convention on interpretation of correlations by Cohen [50] and Durlak [51], we classify the effect size of convenience on satisfaction as middle (0.36) for e-commerce and large (0.52) for voice commerce. Although transaction process efficiency significantly influences satisfaction in voice commerce, it does not show significance in e-commerce. The results for the rest of the constructs were not statistically significant. Recommendation complexity influences e-commerce satisfaction positively. The effect of recommendation personalization on voice commerce satisfaction has an effect size of 0.14 and a p-value of 0.13.

Table 4. Predictors for satisfaction in e-commerce and voice commerce

Path	*Estimate*	*Beta*	*p-value*
EC ↑ECS	0.52	0.36	<0.001***
ETP ↑ ECS	0.04	0.02	0.793
ERP ↑ ECS	0.10	0.09	0.355
ERC ↑ ECS	0.13	0.12	0.228
VC ↑ VCS	0.68	0.52	<0.001***
VTP ↑ VCS	0.34	0.18	0.01*
VRP ↑ VCS	0.16	0.14	0.13
VRC ↓ VCS	-0.01	-0.01	0.896

Significance levels: * p<0.05, ** p<0.01, *** p<0.001; positive effect ↑, negative effect ↓

Table 5 shows the results for the comparative hypotheses in this approach. To assess hypotheses H1-H4, we first performed a t-test to analyze statistically significant differences between datasets. The test resulted in significant difference for all constructs. We then compared the respective path coefficients (Beta) and noted the absolute numerical difference (Delta). Whenever the difference exceeded 0.10, the hypothesis is considered as supported. Convenience significantly influences both e-commerce and voice commerce satisfaction, but clearly does so more in voice commerce. Transaction process efficiency also presents a more sizable effect for voice commerce, as predicted. However, the results present the issue that this construct significantly influences satisfaction only in voice commerce. This indicates that the concept is only relevant (or only valid) in voice commerce. We also compared coefficients that were not found to be significant. The results do not support a difference between recommendation personalization for voice and e-commerce, as the numerical delta is only 0.05. The assessment of recommendation personalization shows a delta of 0.13. Thus, both effects are not significant.

Table 5. Comparative hypotheses results

Hypothesis	*Description*	*Beta*	*Delta*	*t-test*	*Conclusion*
H1	VRC < ERC	-0.01 vs. 0.12	0.13	-5.99***	Not supported
H2	VRP > ERP	0.14 vs 0.09	0.05	-5.15***	Not supported
H3	VC > EC	0.52*** vs. 0.36***	0.16	22.55***	Supported
H4	VTP > ETP	0.18* vs. 0.02	0.16	20.55***	Not supported

Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$; Delta = numerical difference between standardized beta coefficients

5 Discussion

The first objective of this study was to identify and understand factors of customer satisfaction for e-commerce and voice commerce systems. The second objective was to compare these effects between these two channels of e-commerce. We conducted a survey to test the research models. The results confirm one out of four hypotheses and provide some support for the conceptual models. They particularly show that convenience significantly influences customer satisfaction in both e-commerce and voice commerce, and that the effect is in fact larger in voice commerce. Results also show that transaction process efficiency, in terms of overall process speed and number of process steps, significantly influences voice commerce customer satisfaction. This was explained by higher efficiency expectations through increased efficiency of the speech interface. Further the results inferred that users prefer to browse and take their time using the e-commerce channel when compared to voice commerce. According to the results, complexity, extent and degree of detail of recommendation presentation as well as personalization of recommendations do not have a significant effect on satisfaction.

When designing voice commerce applications, developers and designers should keep in mind that convenience and efficiency expectations are higher than those towards e-commerce systems. This may lead to the following design choices: 1) The voice commerce system features increased ease of use and effortlessness over comparable e-commerce systems. 2) The process of searching and buying is designed to be as quick as possible; there are neither detours nor long dialog stages. 3) The number of steps in the process is limited to a necessary minimum. Time intensive input of address or payment data should be omitted. 4) Since a significant number of my sample uses voice commerce on mobile phones, designers should think about creating voice commerce systems for these platforms. There, visual output could be added to increase media richness and usability.

This is the first study to investigate voice commerce customer satisfaction predictors and comparing these with those of e-commerce, which adds knowledge to academic literature and will improve the understanding of the relationships between these system types. It is also one of few studies to compare two structural equation models to assess comparative hypotheses. This approach should increase reliability, because the same participants provide their input on both models.

6 Limitations and Future Research

This study is subject to several limitations, such as sample selection. We collected data from Amazon MTurk, and so reached mainly young users with high IT affinity. However, it is not representative of the general population of any country [52]. Additionally, only US users participated in the survey. While this was motivated by the higher diffusion rate of voice commerce in the US and the large absolute population size, it presents a limitation when it comes to transferability and generalization of the results.

Because voice commerce is an area currently evolving, many opportunities for future research arise. A future study could try to ascertain data from voice-exclusive scenarios for clearer insights into its intricacies and avoid intermixture systems that combine voice and visual interfaces. If however the trend of these systems gains more significance, research should focus on this area.

A dedicated, detailed study to investigate how product complexity interacts with customer buying behavior and decisions, akin to Maity and Dass [23] investigations on this topic in e-commerce, m-commerce and in-store purchasing, could generate insights on how consumers handle search and experience respectively low and high involvement goods in voice commerce. A study covering detailed customer preferences for each channel could shed light to this and similar questions, for example whether repeat purchases are more likely to take place via voice commerce and whether products bought are predominantly of low-complexity as well as which factors generally influence customers in their decision to use voice commerce over other channels.

A number of assumptions concerning recommendations could be assessed more effectively with local, hands-on laboratory settings, especially those motivated by media richness and cognitive overload. For example, in an experiment where participants actually experience the difference between very long and detailed and very short product descriptions and possible cognitive overload, results may be much more distinct than in a self-administered survey. Alternatively, real-life e-commerce environments present opportunities for experiments using A-B testing.

References

1. Broder, A.: A Call to Arms: Embrace Assistive AI Systems! In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. ACM Press, New York, NY (2018)
2. Moar, J.: Digital Voice Assistants: Platforms, Revenues & Opportunities. Juniper Research (2017)
3. Luger, E., Sellen, A.: "Like Having a Really Bad PA": The Gulf betweenUser Expectation and Experience of Conversational Agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5286–5297. ACM Press, New York, USA (2016)
4. Chopra, S., Chivukula, S.: My Phone Assistant Should Know I Am an Indian – Influencing Factors for Adoption of Assistive Agents. In: 19th International Conference on Human-Computer Interaction with Mobile Devices and Services. pp. 1–8 (2017)
5. Kwon, O.B., Sadeh, N.: Applying case-based reasoning and multiagent intelligent system to context-aware comparative shopping. Decis. Support Syst. 37, 199–213 (2004)
6. Galanxhi, H., Nah, F.F.: U-commerce: emerging trends and research issues. Ind. Manag. Data Syst. 104, 744–755 (2004)
7. Brandtzaeg, P.B.: Why People Use Chatbots. In: 4th International Conference on Internet Science (Lecture Notes in Computer Science). pp. 377–392. Springer International Publishing, Thessaloniki (2017)
8. Srinivasan, S.S., Anderson, R., Ponnavolu, K.: Customer loyalty in e-commerce: An exploration of its antecedents and consequences. J. Retail. 78, 41–50 (2002)
9. Yang, S.: Understanding B2B customer loyalty in the mobile telecommunication industry: a look at dedication and constraint. J. Bus. Ind. Mark. 30, 117–128 (2015)
10. Peng, M., Qin, Y., Tang, C., Deng, X.: An E-Commerce Customer Service Robot Based on Intention Recognition Model. J. Electron. Commer. Organ. 14, 34–44 (2016)
11. Li, X., Hsieh, J.P.A., Rai, A.: Motivational differences across post-acceptance information system usage behaviors: An investigation in the business intelligence systems context. Inf. Syst. Res. 24, 659–682 (2013)
12. Cao, Y., Lu, Y., Gupta, S., Yang, S.: The effects of differences between e–commerce and m–commerce on the consumers' usage transfer from online to mobile channel. Int. J. Mob. Commun. 13, 51–70 (2015)
13. Ben Mimoun, M.S., Poncin, I., Garnier, M.: Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics. Inf. Manag. 54, 545–559 (2017)
14. Mahmood, T., Mujtaba, G., Venturini, A.: Dynamic personalization in conversational recommender systems. Inf. Syst. E-bus. Manag. 12, 213–238 (2014)
15. Shawar, B.A., Atwell, E.S.: Using corpora in machine-learning chatbot systems. Int. J. Corpus Linguist. 10, 489–516 (2005)
16. AbuShawar, B., Atwell, E.: Usefulness, localizability, humanness, and language-benefit: additional evaluation criteria for natural language dialogue systems. Int. J. Speech Technol. 19, 373–383 (2016)

17. Horzyk, A., Magierski, S., Mklaszewski, G.: An Intelligent Internet Shop-Assistant Recognizing a Customer Personality for Improving Man-Machine Interactions. Recent Adv. Intell. Inf. Syst. (2009)
18. Qiu, L., Benbasat, I.: Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. J. Manag. Inf. Syst. 25, 145–182 (2009)
19. Christakopoulou, K. Radlinski, F., Hofmann, K.: Towards Conversational Recommender Systems. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2016)
20. Liang, T.-P., Lai, H.-J., Ku, Y.-C.: Personalized Content Recommendation and User Satisfaction: Theoretical Synthesis and Empirical Find. J. Manag. Inf. Syst. 23, 45–70 (2006)
21. Xiao, B., Benbasat, I.: E-Commerce Product Recommendation Agent Use, Characteristics, and Impact. MIS Q. 24, 245–276 (2007)
22. Davern, M., Shaft, T., Te 'eni, D.: Cognition Matters: Enduring Questions in Cognitive IS Research. J. Assoc. Inf. Syst. 13, 273–314 (2012)
23. Maity, M., Dass, M.: Consumer decision-making across modern and traditional channels: E-commerce, m-commerce, in-store. Decis. Support Syst. 61, 34–46 (2014)
24. Hong, W., Thong, J.Y.L., Tam, K.Y.: Designing product listing pages on e-commerce websites: An examination of presentation mode and information format. Int. J. Hum. Comput. Stud. 61, 481–503 (2004)
25. Hostler, R.E., Yoon, V.Y., Guimaraes, T.: Recommendation agent impact on consumer online shopping: The Movie Magic case study. Expert Syst. Appl. 39, 2989–2999 (2012)
26. Li, S., Karahanna, E.: Peer-based recommendations in online B2C ecommerce: Comparing collaborative personalization and social network-based personalization. In: Proceedings of the Annual Hawaii International Conference on System Sciences. pp. 733–742 (2012)
27. Jiang, Y., Shang, J., Liu, Y.: Maximizing customer satisfaction through an online recommendation system: A novel associative classification model. Decis. Support Syst. 48, 470–479 (2010)
28. Komiak, S.Y.X., Benbasat, I.: The Effects of Personalization and Familiarity on The Effects of Personalization Trust and Adoption of Recommendation. MIS Q. 30, 941–960 (2006)
29. Häubl, G., Murray, K.B.: Double Agents: Assessing the Role of Electronic Product Recommendation Systems. MIT Sloan Manag. Rev. 47, (2006)
30. Agarwal, V., Zia Khan, O., Sarikaya, R.: Remembering What You Said: Semantic Personalized Memory for Personal Digital Assistants. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 5835–5839 (2017)
31. Chaparro-Peláez, J., Agudo-Peregrina, A. F. Pascual-Miguel, F.J.: Conjoint analysis of drivers and inhibitors of e-commerce adoption. J. Bus. Res. 69, 1277–1282 (2016)
32. Wang, R.J.H., Malthouse, E. C. Krishnamurthi, L.: On the Go: How Mobile Shopping Affects Customer Purchase Behavior. J. Retail. 91, 217–234 (2015)
33. Choi, J., Seol, H., Lee, S., Cho, H., Park, Y.: Customer satisfaction factors of mobile commerce in Korea. Internet Res. 18, 313–335 (2008)
34. Chai, J.Y., Budzikowska, M. Horvath, V., Nicolov, N., Kambhatla, N., Zadrozny, W., Stys, M., Kambhatla, N., Zadrozny, W., Melville, P.: Natural Language Sales Assistant-A Web-Based Dialog System for Online Sales. In: Proceedings of the Thirteenth Innovative Applications of Artificial Intelligence Conference. pp. 19–26. , Seattle, Washington (2001)
35. Saliba, A.J.: Auditory-visual integration in sound localisation. University of Essex (2001)

Business Process Compliance and Blockchain: How Does the Ethereum Blockchain Address Challenges of Business Process Compliance?

Anja Meironke[1], Tobias Seyffarth[1], and Johannes Damarowsky[1]

[1]Martin Luther University Halle-Wittenberg, Chair of Information Management, Halle (Saale), Germany
{anjameironke}@gmail.com
{tobias.seyffarth, johannes.damarowsky}@wiwi.uni-halle.de

Abstract. Second generation blockchain technologies such as Ethereum can be used not only for financial transactions but also for cross-organizational processes, for applications in the pharmaceutical industry and even in the field of Business Process Compliance (BPC). However, there are many challenges in the field of BPC. Thus, we raised the following research question: How does the Ethereum blockchain address challenges of BPC? To answer this question, we conducted a structured literature review to identify challenges in BPC as well as features of the Ethereum blockchain that may solve the selected BPC challenges. As a result, we identified 21 BPC challenges and categorized these into legal, organizational, human-centered, technical and economic challenges. We found that the technical and organizational BPC challenges were those that Ethereum could best solve, while human-centered challenges could be less well addressed. Furthermore, the implementation of the Ethereum blockchain leads to additional challenges, such as the immutability of illegal content within the Ethereum blockchain or the error-proneness and zero-defect tolerance of smart contracts.

Keywords: Blockchain, Ethereum, Business Process Compliance, Challenges

1 Introduction

In the time of digital innovation, advanced technologies are emerging and changing the way business is done, especially business between organizations. One of the recent technologies which is said to be disruptive in nature is blockchain [1–3]. Currently, the most prominent blockchain technology is Bitcoin, due to its correlated and identically named cryptocurrency [2, 4]. However, modern blockchain deployments such as Ethereum provide even more elaborate functionalities and, therefore, have also gained attention. Ethereum reaches beyond digital currencies, enabling applications in cross-organizational business processes, logistics and pharma [5–7]. Another potential area of application is in Business Process Compliance (BPC), which denotes the execution of business processes in adherence to compliance requirements such as laws or contracts [8]. However, there are various challenges within BPC such as the complexity

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

of business and compliance processes, the high administrative efforts required to ensure compliance in business processes and the lack of automation and standardization to adequately support BPC [9, 10]. Whereas research and earlier blockchain applications mainly concentrated on financial applications and use cases in the field of cross-organizational business process execution [6, 7, 11], further applications are of interest as well. Therefore, our goal is to answer the following research question: *How does the Ethereum blockchain address challenges of BPC?*

The remainder of this paper is organized as follows. In Section 2, we describe the applied methodology to answer our research question. In Section 3, we briefly highlight the technological underpinnings of the blockchain deployment Ethereum and its main features. Section 4 provides an overview of the major challenges within the field of BPC. In Section 5, we map the features of Ethereum to the different BPC challenges and discuss the potential of the findings derived. Finally, Section 6 concludes the paper.

2 Methodology

We performed a three-step approach to answer our research question. First, we worked out the features of the Ethereum blockchain. Second, we identified challenges in the field of BPC, which were then mapped to the Ethereum features in the third step.
We conducted a literature review according to vom Brocke et al. [12] to identify relevant literature for each step. Table 1 shows the applied search terms within each database, the initial hits and the numbers of selected and relevant papers.

Table 1. Literature search

Step	Search term	Database	Hits	Selected	Relevant
1	ethereum AND (functionality OR "mode of operation" OR "way of functioning")	Google Scholar	211	6	5
2	("business process compliance" OR BPC) AND (challenge OR lifecycle)	Google Scholar	Aborted after 610	80	17
	"Business Process Compliance" challenge	KVK	104	12	12
3	("business process compliance" OR BPC") AND (blockchain OR ethereum)	Google Scholar	38	9	1
	“business process compliance” ethereum	KVK	36	28	17
Backward search					19
Sum					71

The literature search was performed using Google Scholar and Karlsruher Virtueller Katalog (KVK), which includes the following databases: GBV, SWB, BVB, HBZ, HEBIS, KOBV, DNB, StaBi Berlin and Worldcat. In Google Scholar we searched within the general search field, while we used the keyword search within KVK, which only allowed restricted search strings and no connectors. The initial hits were selected based on their titles and abstracts. In the next step, the papers' abstracts and several papers in full were read to identify relevant publications. We also conducted a backward search, which led to 71 relevant papers that were considered to help answer the research question.

3 Fundamentals of Blockchain and Ethereum

Blockchain is the technology that supports Ethereum [4]. The Ethereum blockchain is a public and distributed ledger, which stores all of the transactions occurring within the Ethereum network. Transactions are processed between different accounts. Every party of the blockchain network can create any number of accounts without restrictions and third party authentications. The creation of accounts and the authorization of transactions are based on asymmetric cryptographic mechanisms. When creating an account, a public and a private key are generated whereby the account address is derived from the public key, which guarantees a certain degree of anonymity. The sender authorizes the transaction using the account's private key. Processed transactions are visible to every party of the network. Due to the implemented consensus mechanism, the blockchain technology is tamper-proof, without any need for a trusted third party (e.g. a bank or a notary) to avoid double spending.

One consensus mechanism is proof-of-work (PoW), whereby so-called miners have to solve a puzzle using cryptographic methods. Next, new transactions are published in the blockchain network and are approved by the miners. Miners summarize a number of transactions to blocks, validate their signature and the transaction nonce and check that the sender's account balance covers the amount, including the fees to execute the transaction. Then a hash is computed over the current and previous block of transactions and their metadata. A valid hash must correspond to a certain pattern (for example, in the case of Ethereum, the hash must be below a certain threshold). In order to achieve this, a nonce is added as a further input of the hash function. It is not possible to simply compute this nonce. Nonces have to be tested randomly. After finding a matching nonce, the blocks are chained together to form a blockchain. Finally, the miner who first finds a valid nonce is awarded with an amount of new crypto-coins and all transactions fees of the respective block. Based on chaining blocks by means of their computed hash values, a manipulation can be easily detected by recalculating the hashes of two linked blocks and their nonce.

Additionally, there are further design parameters for blockchain solutions. Blockchains can be differentiated between public and private and between permission-less and permissioned blockchains. In a public blockchain, everyone can be part of the blockchain network, whereas private blockchain access is only granted to dedicated participants. Within a permission-less network, everyone can approve new blocks, while in a permissioned network only certain parties are allowed to [13]. By default, Ethereum represents a public, permission-less blockchain [5].

Finally, two generations of blockchains are discussed in the literature. The first generation of blockchain is able to transfer tokens (e.g. Bitcoins) between nodes. The second generation, to which Ethereum belongs, allows more elaborate bytecode to run on top of the blockchain, which is denoted as so-called smart contracts. A smart contract is a user-defined program executed within the blockchain network [14]. In Ethereum, smart contracts are written in Solidity and executed within its execution environment, the Ethereum Virtual Machine. They can be executed automatically and allow a flexible adaption of the blockchain technology to other fields of application [5]. Based on the five technical underpinnings explained above, we derived fourteen main features of Ethereum in Figure 1.

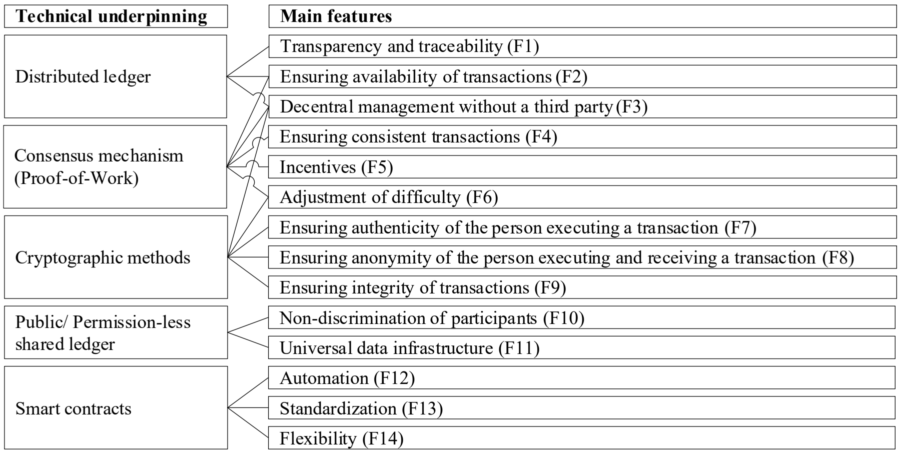


Figure 1. Technical underpinnings and main features of the Ethereum blockchain

4 Challenges in Business Process Compliance

The successful management of BPC is associated with various challenges. We classified these challenges into groups, including legal, organizational, human-centered, technical and economic challenges. Table 2 shows an excerpt of the concept matrix. The entire concept matrix can be found at: `https://bit.ly/2A3O35O`.

Table 2. Concept matrix of BPC challenges (H: Human-centered | T: Technical)

		[15]	[16]	[17]	…
H	Lack of awareness and acceptance (C_H 1)	x	x		
	Conscious or unconscious misconduct (C_H 2)		x		
	…				
T	Technical support and automation of BPC (C_T 1)		x		
	Complex IT architecture and low system integration (C_T 2)		x		
	Proprietary service providers and centralized services (C_T 3)		x		
	Data security and data privacy (C_T 4)			x	
…	…				

Legal challenges include the complexity of the compliance requirements that must be taken into account [8, 10]. A compliance requirement is an assertion resulting from the interpretation of compliance sources, such as laws, norms, standards or internal policies [18]. Compliance requirements change rapidly, which requires a continuous adaptation of related business processes [9, 19]. Moreover, the vagueness of compliance requirements and differing interpretations by the stakeholders make it difficult to interpret compliance requirements unambiguously [15, 20].

Further, organizations are often confronted with **organizational challenges**. In addition to different organizational structures, the complexity and multitude of business and compliance processes can turn BPC management into a difficult task [9, 10]. Organizations cooperate and interact on an increasingly global level [21]. Therefore, not only must internal business processes be checked for compliance, but entire cross-organizational business processes of the involved partner organizations must also be checked [21, 22]. In these cases, BPC management is difficult and time-consuming, especially because different legislation may be applicable for each country and the stakeholders involved may pursue different objectives. Often there is a problem of trust concerning the exchange of information between different organizations [13, 22–24]. Further organizational BPC challenges comprise the modelling and verification of compliance rules and their integration into the corresponding business processes [25]. Additional organizational challenges include the monitoring of compliance requirements and their operationalization in business processes, for example through compliance processes [19], the associated documentation of compliance process execution and verification, as well as adequate transparency and traceability. Business processes and their instances must be checked for compliance during and after runtime by either external auditors or internal evaluations by means of reports or log files [17, 25, 26]. For this purpose, documented evidence must be tamper-proof, available and easily accessible to the parties involved [13, 16]. In addition, the traceability of compliance requirements back to their relevant business process models and vice versa is often inadequately designed, meaning changes of compliance requirements and business processes cannot be implemented adequately, due to a lack of referential clarity [16].

Closely linked to organizational challenges are **human-centered challenges**. The lack of stakeholder acceptance and awareness, combined with insufficient communication and awareness measures on the part of management, impede the successful implementation of necessary BPC measures. Moreover, a successful BPC is limited by either conscious or unconscious misconduct due to deficiencies in knowledge or by the deliberate violation of compliance requirements resulting from malicious intent, a lack of motivation or fear [16].

Technical challenges generally reside in the low level of automation. There are often manually performed BPC tasks, such as the modeling of compliant business processes and the verification and monitoring of business processes according to their compliance requirements, which are not only time-consuming but also error-prone [9, 10, 13, 16, 27]. Moreover, not all tasks within the BPC management lifecycle [28] can be automated; for example the generation of compliance requirements deduced from different compliance sources requires a broad understanding and complex, strategic thinking [16]. Additionally, the management of information technology (IT) architectures is also challenging because these are usually heterogeneous, distributed, isolated and mutually incompatible [15]. Therefore, the integration of common tools to support BPC across several organizations and the reduction of parallel IT systems, as well as redundant and inconsistent data fragments, is a challenging task [29]. Above all, the implementation of tools to support BPC often leads to dependence on the corresponding service provider in terms of financial aspects, data security and data privacy [16]. Moreover, sensitive data of an organization is to be treated confidentially and therefore must be protected against unauthorized access and manipulation [13, 17, 24]. Additionally, a service provider may become a single-point-of-failure, if used as the sole provider for data processing [27]. At the same time, successful BPC management must be able to provide information at all times and therefore requires consistent information that is available and documented in a comprehensible manner [30]. Consequently, the simultaneous protection of privacy and availability of BPC-relevant information are often conflicting objectives [16, 17].

Economic challenges include, among other things, a lack of cost and resource efficiency. Companies and organizations have to face increased compliance, IT and staff-related expenses for the management, evaluation and adjustment of complex compliance requirements and business processes or as a result of legal claims for damages due to uncovered compliance violations [8, 21, 31]. Additionally, due to redundant processes, inconsistent data and a low degree of automation, BPC tasks are often inefficient and time consuming [10, 13]. The lack of adequate methods and indicators to analyze the efficiency and effectiveness of BPC so as to assess their cost-benefit ratio constitutes another difficulty [16, 31]. Finally, poor standardization as well as the ad hoc-oriented tasks of the BPC lifecycle make it difficult to optimize business and compliance processes or to react appropriately to compliance deviations [16].

5 How Does Ethereum Address BPC Challenges?

In the previous sections, we focused on the technical underpinnings and main features of Ethereum and on the identified BPC challenges. In the upcoming section, we answer the research question by explaining how Ethereum can potentially address these BPC challenges. For this purpose, the main features of Ethereum were mapped to the identified BPC challenges. The classification of the results is based on our own assessment and the findings of the literature analysis. Table 3 shows an excerpt of our classification. The complete classification including a short explanation and literature references can be found at: `https://bit.ly/2ygavHk.`

Table 3. Classification of Ethereum features referred to BPC challenges (excerpt)

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14
C_H 1			x									x	x	x
C_H 2	x	x						x	x			x	x	
…														
C_T 1	x											x		x
C_T 2	x		x	x						x	x			x
C_T 3		x	x						x		x			
C_T 4	x	x	x		x		x	x	x					
…														

In terms of **legal BPC challenges**, there is low potential for improvement by the features of Ethereum. According to [32], it is conceivable to store compliance requirements, such as laws and standards, as independent objects of a smart contract within the Ethereum blockchain to make them tamper-proof and publicly accessible. Thus, compliance and business process managers can be informed automatically if changes in the status of a referred requirement occur.

Ethereum offers promising opportunities to address **organizational BPC challenges**, such as the modeling of business processes according to their compliance requirements and the monitoring of compliance requirements and verifying their documentation, transparency and traceability. The decentralized and distributed structure of the Ethereum blockchain supports the streamlining of business and compliance processes, especially with regard to cross-organizational business processes. Accordingly, certain organizational units or trusted third parties that are in charge of monitoring BPC can be omitted (e.g. the central banks for foreign transfers), since transactions may be carried out directly between the transaction partners [7, 33]. Using the Ethereum blockchain as a universal data infrastructure for accessing and coordinating shared data and processes between different stakeholders also reduces data traffic and complex procedures necessary for reconciling, managing and controlling data and processes [34, 35]. In addition, Ethereum contributes to the modelling and verification of compliance rules and their integration into business processes. Logics and rules associated with business and compliance processes can be expressed and automatically checked with the help of smart contracts. However, business and compliance processes as well as their associated rules cannot always be translated seamlessly from natural language into program code [36]. Additionally, due to the overly domain-specific expertise of programmers or legal experts, smart contracts may be programmed inaccurately and might not reflect the underlying rules and process models as intended [36, 37].

Besides supporting BPC-modelling, Ethereum also offers various options for monitoring, verification and transparent documentation within the context of BPC. With the help of smart contracts, a large part of BPC tasks like monitoring and verification can be automated and outsourced to Ethereum, thus reducing manual effort and various process activities [1, 11, 13]. Applying an if-then pattern, smart contracts also automatically enforce compliance with defined rules, with the result that violations cannot occur in the first place, making a permanent monitoring unnecessary. An illustrative example would be a locked leased car that can only be opened by means of a digital key after payment was made [33].

Additionally, the connection to or general accessibility of the blockchain as a public and shared ledger enables extended transparency towards internal and external compliance units and supervisory authorities [7]. Transactions that are documented in the Ethereum blockchain in a verifiable and tamper-proof manner can be verified by internal and external auditors and legislators [2, 13, 17, 38]. In addition, specific verifiable documents and records (e.g. certificates, contracts) or material assets (for example motor vehicles, land) can be stored as digital assets. If their status changes, for example in the case of value transfers and changes of ownership, changes can be documented in a tamper-proof way [7, 11, 13, 39].

However, relevant process participants, such as supervisory authorities or employees, have to be an integral part of the Ethereum network to provide access to information on transactions [33, 37]. Moreover, illegal contents (e.g. child pornography or unauthorized personal data) are also problematic, since they cannot be simply removed from the blockchain and therefore risk making the entire blockchain illegal [3].

In contrast to the organizational aspects mentioned above, Ethereum has a moderate potential for solving **human-centered BPC challenges**. Ethereum shows potential in terms of reducing conscious (e.g. fraud and deliberate disregard of rules) or unconscious (e.g. lack of knowledge) misconduct. On the one hand, knowledge deficits, errors or deviations can be diminished by mapping compliance requirements, compliance processes or entire business processes as smart contracts for (partial) automation so that they are prevented from circumvention [33, 37, 38]. On the other hand, the manipulation of transaction data causes high costs due to the enormous computing effort necessary to find a valid nonce. Furthermore, fraud and erroneous practices are immediately recognizable due to real-time transaction processing and transparent documentation in the public ledger, so that problems like payment defaults or insurance fraud can be avoided [1, 7, 37].

Ethereum also contributes significantly to addressing **technical BPC challenges**. This particularly concerns aspects of automation, IT architecture, data security and data privacy. As already mentioned, entire business and compliance processes can be (partially) automated with the help of smart contracts [3, 33, 37, 39]. When defined events act as triggers or when pre-defined conditions are met, the smart contract code is validated and executed by the miner according to if-then patterns and, finally, documented in the Ethereum blockchain [32, 36, 38, 40]. However, with regard to deviations from defined rules or algorithms during execution, [33] considers the zero-defect tolerance of smart contracts to be problematic. What happens, for example, in terms of a package delivery, if the order is delivered later than stipulated in the smart contract, but has nevertheless been delivered?

In addition, the Ethereum blockchain can be used as a universal infrastructure to connect and coordinate different stakeholders and organizations, which is often a serious challenge, especially with regard to the modeling, verification and monitoring of compliant business processes within BPC [11, 13, 33]. Redundant and non-compliant business processes and data management may be reduced, as the same business processes and data basis can be used equally by the parties involved [10, 13, 33]. The question that arises at this point is to what extent a public blockchain is suitable for making (partly) sensitive data from cross-organizational business processes generally accessible and public [33].

Nevertheless, with regard to aspects of data security, Ethereum has a particularly high potential due to its inherent properties. In contrast to client-server architectures, the decentralized structure of the Ethereum blockchain and its redundant data distribution ensure a high degree of reliability [7, 33]. If a party of the blockchain network fails due to technical faults or manipulation, all other parties have a local copy, ensuring that the data remains available [2]. Moreover, Ethereum offers an additional security mechanism by charging transaction fees for code execution, which ensures that transactions are reversed when their limit is exceeded. Thus, failures due to maliciously built-in loops causing a program code to run infinitely are prevented [40].

The most important feature of the Ethereum blockchain is the guarantee of integrity by means of cryptographic hashing and linking of the data blocks, as described in

Section 3. Due to high energy costs incurred for recalculation, attacks would be unprofitable for attackers [1, 2]. In addition, manipulated data and attempted fraud would be detectable not only in the context of consensus checks, but also after publishing the data in the Ethereum blockchain [24, 33, 39]. Furthermore, the authenticity and anonymity of the transaction participants can be guaranteed by the use of digital signatures and anonymous account addresses (see Section 3). [1, 10, 13, 39].

According to [41], however, there are vulnerabilities regarding long-term data security. Due to improved computing power and mathematical advances, it is doubtful whether current cryptographic hash methods will maintain their current security level in the foreseeable future. Then the data within Ethereum could be compromised retroactively [23]. Besides, users do not know where their data is stored or to what extent it is processed [42].

Finally, Ethereum offers potential approaches for the solution of **economic BPC challenges,** such as suboptimal cost and resource efficiency, measurability and standardization. Thus, the automation of compliance processes and business processes and their streamlining, by eliminating unnecessary process activities and intermediaries, enables a better cost and resource efficiency, especially on a cross-organizational level [6, 7, 11]. Cost and resource efficiency benefits can also be realized using the Ethereum blockchain as a universal data infrastructure for managing shared data, since multiple data checking by the individual transaction units might become obsolete. This can reduce inconsistent, redundant information as well as the associated work effort, since the Ethereum blockchain is turned into a single-point-of-truth [2, 32, 33].

Interesting approaches also arise in terms of the measurability of BPC efficiency and effectiveness. Due to the public and shared nature, it is possible to implement comprehensive methods for a better monitoring of transaction data. Thus, BPC-relevant process and transaction data can be collected and evaluated in real time [10, 13, 24]. Since transaction fees have to be paid and compliance processes can be partially automated by smart contracts, BPC costs can be determined on process activity level [43].

However, the high energy consumption for performing the PoW and the local memory requirements for redundant blockchain replications are considered to be critical [44]. All in all, Ethereum offers comprehensive possibilities to address different BPC challenges. Nevertheless, resulting risks must be considered as well.

6 Conclusion

The second generation of Blockchain technologies, such as Ethereum, allows not only for the processing of financial transactions [11], but also for elaborated applications in various fields, such as BPC. Since there are a lot of BPC challenges (e.g. [9, 10]), we raised the following research question: How does the Ethereum blockchain address challenges of BPC? By conducting a literature review, we identified 14 main features of Ethereum. Furthermore, we identified 21 BPC challenges and categorized them according to legal, organizational, human-centered, technical and economic challenges.

A large portion of the challenges is of organizational and technical nature. It has been shown that the main contribution of the Ethereum blockchain is to solve technical and organizational challenges, whereas human-centered challenges are less solvable. The following main features are of utmost importance to meeting BPC challenges: automation with the help of smart contracts; the transparent design and traceability of the public ledger; and the possibility of using Ethereum as a universal data infrastructure. However, the use of the blockchain also results in new risks, such as the immutability of illegal content or the error-proneness of smart contracts due to a lack of knowledge or their zero-defect tolerance during execution [3, 33, 36, 37].

A well-known shortcoming of any literature review is the fact that it is not possible to consider all relevant work. However, by documenting the literature research according to vom Brocke et al. [12], comprehensibility in the development of arguments is provided in a scientific manner. The assignment of BPC challenges and Ethereum properties is not always documented in the literature and therefore results from an argumentative assignment. Additionally, each property of the Ethereum Blockchain in our reference table has an equally weighted potential to meet the respective BPC challenges. Our future research aims to investigate these potentials explicitly, according to their specific impact on solving BPC challenges.

References

1. Aste, T., Tasca, P., Di Matteo, T.: Blockchain Technologies. The Foreseeable Impact on Society and Industry. Computer, 50(9), 18–28 (2017)
2. Asharaf, S., Adarsh, S.: Decentralized computing using blockchain technologies and smart contracts: Emerging research and opportunities. Introduction to Blockchain Technology. IGI Global, Hershey, PA, USA (2017)
3. Ateniese, G., Magri, B., Venturi, D., Andrade, E.: Redactable Blockchain - or - Rewriting History in Bitcoin and Friends. In: IEEE EuroS&P 2017, pp. 111–126
4. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System, https://bitcoin.org/bitcoin.pdf (Accessed: 12.05.2018)
5. Buterin, V.: Ethereum White Paper. A next generation smart contract & decentralized application platform, https://www.weusecoins.com/assets/pdf/library/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf (Accessed: 18.06.2017)
6. Bocek, T., Rodrigues, B.B., Strasser, T., Stiller, B.: Blockchains everywhere - a use-case of blockchains in the pharma supply-chain. In: IFIP/IEEE IM 2017, Lisbon, Portugal, pp. 772–777
7. Deubel, M., Moormann, J., Holotiuk, F.: Nutzung der Blockchain-Technologie in Geschäftsprozessen. In: INFORMATIK 2017, Chemnitz, Germany, pp. 829–842
8. Schäfer, T., Fettke, P., Loos, P.: Towards an Integration of GRC and BPM - Requirements Changes for Compliance Management Caused by Externally Induced Complexity Drivers. In: BPM 2011, Clermont-Ferrand, France, pp. 344–355
9. Becker, J., Delfmann, P., Eggert, M., Schwittay, S.: Generalizability and Applicability of Model-Based Business Process Compliance-Checking Approaches — A State-of-the-Art Analysis and Research Roadmap. Business Research, 5(2), 221–247 (2012)

10. Khan, C., Lewis, A., Rutland, E., Wan, C., Rutter, K., Thompson, C.: A Distributed-Ledger Consortium Model for Collaborative Innovation. Computer, 50(9), 29–37 (2017)
11. Modsching, M., Apfelbacher, A., Horch, J., Kummer, N.: Using a blockchain-based approach to exchange (financial) assets. Journal of digital banking, 2(2), 110–122 (2017)
12. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant. On the importance of rigour in documenting the literature search process. In: ECIS 2009, Verona, Italy, pp. 2206–2217
13. Fridgen, G., Radszuwill, S., Urbach, N., Utz, L.: Cross-Organizational Workflow Management Using Blockchain Technology - Towards Applicability, Auditability, and Automation. In: HICSS-51 2018, Hilton Waikoloa Village, Hawaii, pp. 3507–3516
14. Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., Mendling, J.: Untrusted Business Process Monitoring and Execution Using Blockchain. In: BPM 2016, Rio de Janeiro, Brazil, pp. 329–347
15. Sadiq, S., Governatori, G.: Managing Regulatory Compliance in Business Processes. In: Brocke, J., Rosemann, M. (eds.) Handbook on Business Process Management 2. Strategic Alignment, Governance, People and Culture, pp. 159–175. Springer, Berlin, Heidelberg (2010)
16. Elgammal, A.F.S.A.: Towards a comprehensive framework for business process compliance. Diss. Tilburg University, Netherlands (2012)
17. Kaaniche, N., Laurent, M.: A blockchain-based data usage auditing architecture with enhanced privacy and availability. In: IEEE NCA 2017, Cambridge, MA, USA, pp. 1–5
18. Turetken, O., Elgammal, A., van den Heuvel, W.-J., Papazoglou, M.: Enforcing compliance on business processes through the use of patterns. In: ECIS 2011, Helsinki, Finland. 5.
19. Seyffarth, T., Kühnel, S., Sackmann, S.: A Taxonomy of Compliance Processes for Business Process Compliance. In: BPM 2017, Barcelona, Spain, pp. 71–87
20. Fellmann, M., Zasada, A.: State-of-the-Art of Business Process Compliance Approaches: A Survey. In: ECIS 2014, Tel Aviv, Israel, pp. 1–17
21. Sadiq, S.: A Roadmap for Research in Business Process Compliance. In: BIS 2011, Poznań, Poland, pp. 1–4
22. Knuplesch, D., Reichert, M., Fdhila, W., Rinderle-Ma, S.: On Enabling Compliance of Cross-Organizational Business Processes. In: BPM 2013, Beijing, China, pp. 146–154
23. Zhang, P., Walker, M.A., White, J., Schmidt, D.C., Lenz, G.: Metrics for assessing blockchain-based healthcare decentralized apps. In: IEEE Healthcom 2017, Dalian, China, pp. 1–4
24. Imeri, A., Khadraoui, D.: The Security and Traceability of Shared Information in the Process of Transportation of Dangerous Goods. In: IFIP NTMS 2018, pp. 1–5
25. Rinderle-Ma, S., Thao Ly, L. and Dadam, P.: Aktuelles Schlagwort: Business Process Compliance, https://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.emisa/Downloads/EMISA_aktschlagwort_compliance.pdf (Accessed: 25.02.2018)
26. Reichert, M., Weber, B.: Enabling Flexibility in Process-Aware Information Systems. Challenges, Methods, Technologies. Springer, Heidelberg, Berlin (2012)
27. Matsumoto, S., Reischuk, R.M.: IKP: Turning a PKI Around with Decentralized Automated Incentives. In: IEEE SP 2017, San Jose, CA, USA, pp. 410–426
28. Sackmann, S., Kühnel, S., Seyffarth, T.: Using Business Process Compliance Approaches for Compliance Management with regard to Digitization: Evidence from a Systematic Literature Review. In: BPM 2018, Sydney, Australia, pp. 409–425
29. Brammertz, W., Mendelowitz, A.I.: From digital currencies to digital finance. The case for a smart financial contract standard. The Journal of Risk Finance, 19(1), 76–92 (2018)

30. Kavassalis, P., Stieber, H., Breymann, W., Saxton, K., Gross, F.J.: An innovative RegTech approach to financial risk monitoring and supervisory reporting. The Journal of Risk Finance, 19(1), 39–55 (2018)
31. Kühnel, S.: Toward Cost-Effective Business Process Compliance. A Research Agenda. In: INFORMATIK 2017, Chemnitz, Germany, pp. 2379–2384
32. Yoo, S.: Blockchain based financial case analysis and its implications. APJIE, 11(3), 312–321 (2017)
33. Morabito, V.: Business Innovation Through Blockchain. The B³ Perspective. Springer International Publishing, Cham (2017)
34. Cen, Y., Wang, H., Li, X.: Improving Business Process Interoperability by Shared Ledgers. In: IEEA 2017, Jeju, Republic of Korea, pp. 89–93
35. Kharitonov, A.: A Framework for Strategic Intra- and Inter-Organizational Adoption of the Blockchain Technology. SSRN Electronic Journal, 1–6
36. Frantz, C.K., Nowostawski, M.: From Institutions to Code. Towards Automated Generation of Smart Contracts. In: IEEE FAS*W 2016, Augsburg, Germany, pp. 210–215
37. Clack, C.D., Bakshi, V.A. and Braine, L.: Smart Contract Templates: foundations, design landscape and research directions, https://arxiv.org/pdf/1608.00771v2.pdf (Accessed: 11.05.2018)
38. Reijers, W., O'Brolcháin, F. and Haynes, P.: Governance in Blockchain Technologies & Social Contract Theories, https://ledger.pitt.edu/ojs/index.php/ledger/article/view/62/51 (Accessed: 11.05.2018)
39. Lemieux, V.L.: A typology of blockchain recordkeeping solutions and some reflections on their implications for the future of archival preservation. In: IEEE Big Data 2017, pp. 2271–2278
40. Delmolino, K., Arnett, M., Kosba, A., Miller, A., Shi, E.: Step by Step Towards Creating a Safe Smart Contract: Lessons and Insights from a Cryptocurrency Lab. In: FC 2016, Christ Church, Barbados, pp. 79–94
41. Conte de Leon, D., Stalick, A., Jillepalli, A., Haney, M.: Blockchain: Properties and misconceptions. APJIE, 11(3), 286–300 (2017)
42. Fabiano, N.: Internet of Things and Blockchain. Legal Issues and Privacy. The Challenge for a Privacy Standard. In: IEEE iThings and IEEE GreenCom and IEEE CPSCom and IEEE SmartData, Exeter, UK, pp. 727–734
43. García-Bañuelos, L., Ponomarev, A., Dumas, M., Weber, I.: Optimized Execution of Business Processes on Blockchain. In: BPM 2017, Barcelona, Spain, pp. 130–146
44. Marsal-Llacuna, M.-L., Oliver-Riera, M.: The standards revolution. Who will first put this new kid on the blockchain? In: ITU K 2017, pp. 1–7

Improving Business Model Configuration through a Question-based Approach

Sarah Rübel[1], Adrian Rebmann[1], Andreas Emrich[2,3], Sabine Klein[2,3], and Peter Loos[2,3]

[1] Saarland University, Saarbrücken, Germany
{s9sarueb,s9adrebm}@stud.uni-saarland.de
[2] German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany
{andreas.emrich, sabine.klein, peter.loos}@dfki.de
[3] Institute for Information Systems (IWi), Saarland University, Saarbrücken, Germany
{andreas.emrich, sabine.klein, peter.loos}@iwi.uni-sb.de

Abstract. In the competitive context of agile innovation cycles, it is necessary for companies to construct their business model leading them to a creative strategy innovation. There already are a number of methods to create business models, many of which are also implemented in software. However, these are often unstructured, unguided and static, resulting in diverse and heterogeneous business models. This complicates automated evaluation allowing recommendations. The aim is to develop a question-based tool yielding for comparable and, thus, analyzable business models based on a developed standardized taxonomy. The questions guiding through the configurator were derived from this taxonomy. A tool was developed implementing the question-based concept. User tests were conducted as part of an evaluation showing promising results concerning usability in addition to the already achieved standardization.

Keywords: business model, business model configuration, business model innovation, tool development, question-based design

1 Introduction

1.1 Motivation

Today's development in digitalization and fast innovation shows the necessity for companies to remain profitable and competitive [1]. The constraint to keep up with change pressures companies to rush transformation without really knowing where to begin [2–4]. One significant trend to achieve durable success in the market is to utilize business models as a navigation instrument towards increased profits or finding their niche in the market [5]. They indirectly define the strategy being realized in an organization [6].This explains the necessity of a consulting-based support besides conventional consulting services.

Hence, a faster and simpler way to promote innovation in a complex, uncertain and dynamic environment needs to be available [7, 8]. By describing an organization in a

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

clearly arranged way, business models display weaknesses and potentials [9]. However, even with newly emerging business modelling tools, the construction process can still be considered a self-service, where companies utilize implicit knowledge to depict their organization [10]. The common online business model configurator comprises business model canvases and removable and exchangeable sticky notes [11, 12]. Therein lies the difficulty of business model configuration. Lacking any degree of moderation, users are left to device elements of their own. A question-based instruction of modelling a business can give users momentum to ponder their strategic direction without any refactored domain knowledge and take advantage of already available explicit knowledge [10, 13].

1.2 Objective

The research objective is to improve already existing modelling methods by using a question-based approach to guide the user more efficiently and effectively through the configuration process.

Our central artifact is an online-based business model configurator which should enable companies to describe their business model as completely as possible and on a high enough level to create a certain degree of standardization. For this, the construction can be based on a general process framework reducing the effort for the user. Focus of this paper is not to describe how the developed tool works in detail, but much rather the improvement of business model configuration. Based on a developed questionnaire, the user can fill out their business model by providing input through answering questions [14]. The aspects stem from a generic, high-level business model framework. If companies want to further expand their model, they can add items, which can durably be included in the configurator. The configurator's implementation allows technologies to be matched to certain business model aspects. Ultimately, interdependencies can be detected when a significant number of users is reached.

The question-based configurator should be easy to use and aim at enabling an intuitive, flexible and guided business modelling with a low error ratio. This offers an effective and standardized solution considering more aspects than would be possible if the practitioners fill out an unguided business model canvas. It also has to support the efficient and effective construction of an informative business model with minimal effort. It also enables companies to understand, map and share the business logic [8].

1.3 Methodology & Structure

The derivation of the question-based business model configuration is based on the design science approach [15]. Iterative steps help reevaluate and validate findings at any point during the construction phase [15]. After a research review of business models and existing business modelling tools, the business model framework will be explained, which is based on an extensive literature analysis, which ultimately poses the foundation for the questionnaire [16]. It will utilize the building block system according to the business model canvas from Osterwalder et.al. [5, 8]. As a result, a software product is conceived in form of a business modeling tool [17]. The evaluation is conducted with unrelated test subjects to validate the effectiveness and efficiency of the configurator over an unguided paper-based canvas and post-it approach.

The paper itself is structured as followed: in chapter 2, a general research review of business models and existing modelling tools is given. In section 3, the business model framework and the questionnaire are constructed followed by the proof-of-concept in section 4. Chapter 5 depicts the evaluation and section 6 concludes the paper and gives implications on further research and developments.

2 Related Work

2.1 Business Model

Business models essentially describe how organizations function [18]. They reproduce a generic strategy as a mapped-out list of aspects the organization focuses on in order to remain successful. A digital business organization should be reviewed continuously to ensure its consistency concerning effectivity and efficiency [4]. Thus, a business model is a simplified structured image of reality and consists of connections and elements with characteristics and relations [19, 20]. However, different component compilations exist due to the lack of a uniform understanding of which elements constitute a business model and business models themselves. Depending on the objectives an organization has, the number of potential business models is endless [21]. Therefore, a standardized taxonomy needs to be developed.

As part of a continuous business model improvement, influencing factors such as political, economic, social, technological, legal or ecological changes have to be considered [22]. This enables business model innovation by consciously altering the already existing business model [23]. Hence, a flexible and adaptable way of assembling a business model has to be used to continuously work and improve it. One example of such a tool is the business model canvas by Osterwalder [5].

2.2 Business Modelling Tools

Besides the most frequently used business model canvas [5, 24], a number of other canvas-based business modelling tools can be found online [25]. These tools focus on the visualization of business models, closely resembling the business model canvas, but non-offer any support during the configuration.

As an example of a not question-lead canvas-based configurator, the Canvanizer [11] does not provide any support. It is a static tool to textually describe the user's business model, basically digitalizing the otherwise paper-based canvas approach. The only reference to the required input comes from the title of the building block as determined by Osterwalder [5], but it does not lead through the configuration process.

The only identified semi question-lead canvas-based configurator is the Start-Green [12] sustainable business model configurator. For each building block, different key questions are specified. However, the questions have a strong reference to sustainability aspects, instructing the user to describe their company textually on a very high and abstract level. The user is given the option to declare a question irrelevant and further explanations are provided through keyword examples. Nonetheless, no answers are available towards which the user can be oriented. This prevents standardization and leads to fewer considered aspects then otherwise possible.

3 Concept Development

In the previously mentioned tool survey [25], the business model canvas is the basis of most of the identified construction tools. Since this is also the tool most users are familiar with, it seems appropriate to also base the tool to be introduced in this paper on the business model canvas by Osterwalder. In order to facilitate a certain degree of standardization of the resulting business models, a taxonomy of possible aspects constituting the business model has to be created. Building on an extensive literature analysis, aspects subdividing Osterwalder's building blocks were created, forming an incomplete and not disjoint taxonomy as can be seen in table 1.

Table 1. Extract from the business model taxonomy

Building Blocks	Aspect	Items
Source of Income	Distribution Type	Sales
		Royalties
		License Model
		Leasing, Rental
		…
	Revenue Stream	Core Business
		Diversification
		Large Product Portfolio
	Payment Terms	…
	…	

This is necessary, since the criterion of completeness is unachievable and a termination has to be made after a certain hierarchy level. The aim is to allow users to extent the taxonomy depending on their own requirements, ultimately also lowering the frustration because researchers have not considered all potential responses [26]. Eventually, these aspects will be durably considered in the taxonomy after a certain threshold is crossed. A questionnaire is a formalized set of unambiguous questions translating the researcher's information need to questions, in order to obtain information from applicants [27]. By offering specific response options through a drop-down box, the collection of standardized data is possible. Subsequently, this data can be compared and analyzed. It also enables a faster and more accurate recording of the data. The target is to enable the user to configure their company's business model in its entirety. Hence, the questions themselves and the questionnaire's structure are predefined by the underlying business model taxonomy. The questions are worded around the aspects using natural language to convey further information to the user.

Most questions follow the multiple-choice structure, covering a full range of possible mutually exclusive and collectively exhaustive alternatives. This also eliminates the wording problem, since it is aligned with the taxonomy and the provided answers.

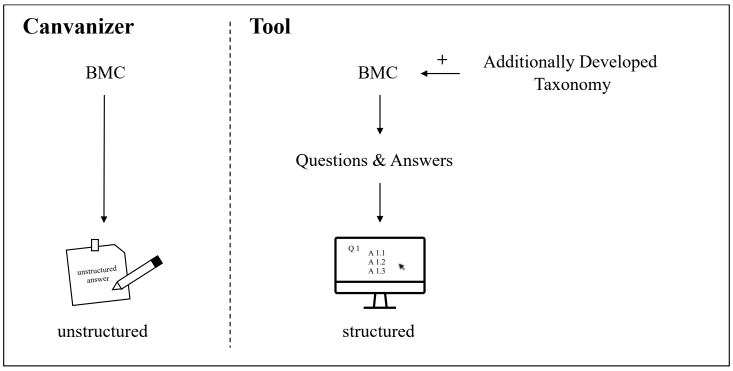


Figure 1. Schematic diagram of the approach

Each selected item is linked to a free text box where they can further explain their selection. For these boxes, information is provided describing the kind of information which can be added in each respective box. This data will only be collected to display in the user's final business model. Since the users are either part of a company's strategy team or, in case of a SME, know the company as a whole, it is ensured that the person using the configurator can actually provide the requested information [28, 29].

Through the question-based approach, refactored domain knowledge is not necessary. By providing answers and giving the option to further expand or explain their selection, the user is also encouraged to be creative and explore different possibilities in the sense of giving more answers than otherwise possible and inspire business model innovation. Whereas, in non-question-based approaches, the level of detail of the resulting business model solely relies on the creativity and experience of the user as can be seen in figure 1. Also, standardization is created allowing the resulting models to be compared and analyzed.

4 Tool Implementation

Based on the conceptualization, an exemplary software prototype was developed that implements the requirements defined in the previous sections. It is built as a modular system as shown in figure 2. The chosen web-based architecture was determined by the framework conditions of the project, the tool is part of. The system consists of two separate user interfaces and a backend. One of the user interfaces offers administrative functionalities. This includes the definition of the business model taxonomy and the creation of the question sequences, the definition of answers to the respective questions and the matching of questions and answers to aspects of the developed business model taxonomy. To do that, the administrator can choose from aspects of the taxonomy or chose his own naming, that can also be linked to a free text field. The question type, i.e. whether only one, multiple or a specific number of answers to a question can be selected.

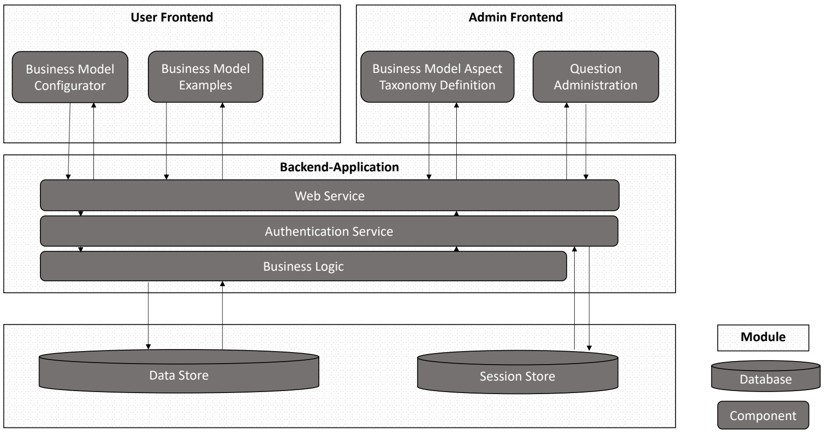


Figure 2. Architecture of the implemented tool

The other UI targets the end user. A user can create, edit and delete business models that are based on the business model canvas. When creating a new business model, the user is first guided through the canvas with the task of answering a number of questions per canvas element. Figure 3 shows an example of such a question as well as the interaction mode via selectable checkboxes and the navigation. Researchers rarely spend sufficient time on the physical layout of their questionnaire, believing that the science lies in the content of the questions and not in such details as the font size or color. Yet empirical studies have repeatedly shown that low response rates are often due to participants being unable to read or follow the questionnaire [28,29].

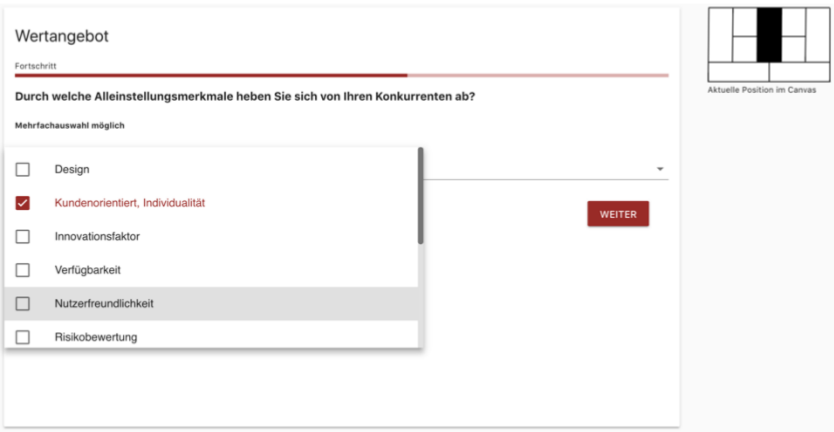


Figure 3. Screenshot of the developed tool's question approach

These aspects are considered in the developed tool, which uses only a small number of UI elements and only a primary and a secondary color. The current position within the canvas as well as the current progress is continuously shown to the user. Navigating between the building blocks of the canvas is also possible. Apart from the actual questions and answers, the aspects are explained using examples. Once the user has answered the questions, the resulting business model is shown in a canvas view.

Figure 4 shows an excerpt of the questions used in the configurator together with their corresponding headline, which are based on the aspects. Answers to the questions are the taxonomy aspects on the lowest hierarchy level.

Building Block	Headline	Question	Type
Customer segment	Market	In which market do you offer your products and services?	Single
	Business relation	With whom does your company maintain business relations?	Multiple choice
	Customer region	Where are your customers settled?	Single
	Industry sector	Which industry sectors do your customers belong to?	Multiple choice
Value Proposition	Products	What kind of products do you offer your customers?	Multiple choice
	Service	What kind of services do you offer your customers?	Multiple choice
	Corporate Responsibility	Through which aspects do you display your sense of responsibility?	Multiple choice
	USP	Through which unique selling proposition do you differentiate yourself from your competitors?	Multiple choice

Figure 4. Excerpt of the developed question catalogue.

This view allows to navigate through the building blocks of the canvas to specify and edit aspects of the created business model with free text elements being offered as shown in figure 5. In this view the user can also decide to revisit some of the questions individually or answer all questions of a building block again. The overview also offers this opportunity and allows the user to download and print the resulting canvas. Since the configurator collects sensitive data, trust in the security of their data needs to be established. This is achieved by implementing a secure login, a session-based authentication and the data being stored in an on-site hosted database. Security aspects are implemented in the backend application. All requests are directed to a web-service that calls an authentication middleware. Valid requests are forwarded to the business logic, which queries a database storing the user specific business model data or if an administrator request arrives taxonomy and question definition data.

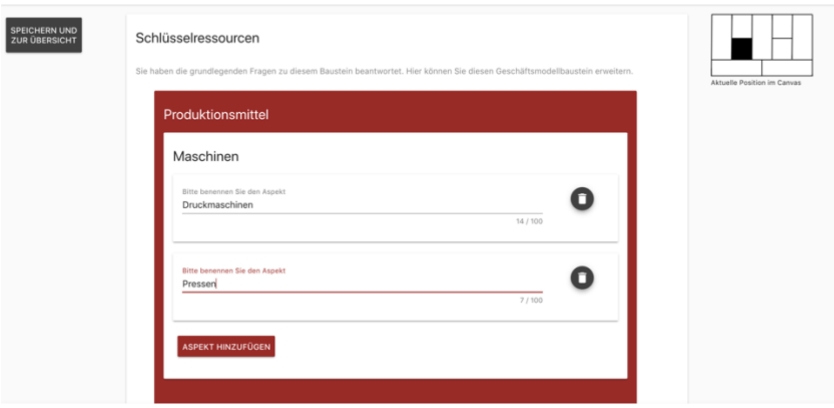


Figure 5. Screenshot of the developed tool’s aspect view

5 Evaluation

5.1 Evaluation Approach

The developed concept and its implementation can be evaluated in various ways. However, a content-based evaluation of the resulting business models is difficult to carry out within the limited time. An evaluation aimed at the comparability of the results could be done with an ontology mapping. However, since the approach developed in this paper is based on a taxonomy, it can be expected in advance that an object of comparison will perform worse in matters of comparability and standardization.

Therefore, the effectiveness of the developed concept will be demonstrated in a case study aiming at the usability evaluation of the implemented prototype. The study design was as follows: a user test was conducted with the following hypothesis: the developed question-based tool is more usable in terms of the NASA TLX score, than non-question-based tools. In the experiment, the system serves as the independent variable with two levels: the implemented prototype, using the concept proposed in this paper and the Canvanizer introduced before, which is representative for non-question-based business model configuration. A semi-question-based tool was not considered, since only one could be found and the target was to evaluate the difference in usability compared to a question-based approach. The dependent variable is the previously mentioned NASA TLX score [32], resulting from an established questionnaire for evaluating mental demand, physical demand, temporal demand, performance, effort and frustration when interacting with a system to accomplish a task. The experiment task consists of the creation of a business model for a well-known company whose name and details are handed to the participants at the beginning of the experiment. Based on the knowledge of said company, the participant is asked to create a business model with the respective tool. As the task of creating a business model is complex, a between-subject approach was used, where each participant is assigned a group. Each group operates on one of the tools. This avoids learning effects or fatigue effects.

Each group had a size of ten participants, each participant was asked to construct a business model on their own. The participants had a general business background. This supports the assumption that the configurator can also be used by people who have no strategic or entrepreneurial prior knowledge. First, the participant was given a short introduction into the concept of business models and was told the company the business model should be created for. After they had finished their task, they were asked to fill out the NASA TLX questionnaire, which asks for the assignment of subjective weights to several evaluation criteria. Finally, they were asked several questions about the system they used.

5.2 Results

The participants were asked to fill out the NASA TXL questionnaire including pairwise comparisons of the evaluation criteria, yielding weighted scores as well as raw scores.

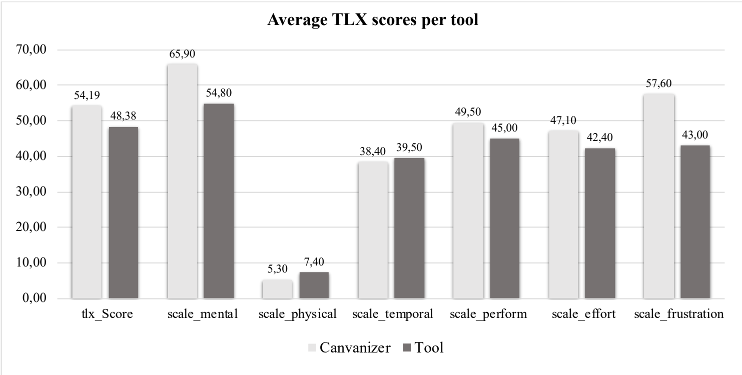


Figure 6. Average TLX scores per tool

Figure 6 shows the average global TLX score and the average scores of the individual criteria per tool. On a scale from 0 to 100, where 100 is the worst possible score. The raw scores are well-suited to evaluate the individual criteria per and across participants to identify general defects. The weighted scores emphasize, which aspects are of central importance for the participant and reveal which criteria contributed most to their workload, when executing the task. Each participant compares the six criteria in pairs. The number of times a criterion is selected represents the weight, which is multiplied by the corresponding raw score and then divided by 15, to get the weighted score. The weighted scores show how much significance the participants assigned to the influence of the criteria to the workload when performing their task.

5.3 Discussion

The results indicate that the participants consider the question-based tool to be significantly less frustrating to use. The improvement compared to the non-question-based tool is clear in the raw scores, but even clearer when the criteria are weighted. The deviations found in the average score were validated and emphasized by the weighted values, because users creating the business model with the non-question-based tool have to ponder about the aspects they want to include. In contrast, the developed tool already gives hints to the expectation of what should be included in a particular building block. This allows for a more creative interaction with additional aspects of innovation. Moreover, the participants found the mental demand to be considerably lower when using the question-based tool, suggesting that the guidance of the user was successful.

Apart from that, the question-based tool performed better when looking at the global TLX score. Except for the mental demand all scores of the question-based tool are below 50 indicating that those criteria are met. Even though on average, the participants, that used our tool, took 33% longer to complete their task, the estimation of temporal effort is only slightly higher. This suggests, that the use of the tool is perceived as engaging by the users.

6 Conclusion & Outlook

The developed tool is a flexible, intuitive and question-based alternative for a standardized business model development, ultimately supporting innovation through an overview-based display of highly comparable business models. It offers an effortless and user-friendly tool for an effective, efficient and guided business model configuration with low error ratio. The resulting business models are homogeneous in terms of detail. This is the case down to a certain hierarchy level, when the taxonomy is terminated for standardization reasons and because completeness cannot be achieved. Evaluation results have shown, that the mental demand is significantly lower in the developed tool than the Canvanizer, implying a success in the developers' effort to make to tool more user friendly. Test subjects stated, that aspects were offered which the user would not have considered otherwise and expressed their content with the instructions and the resulting business models. Furthermore, the evaluation showed, that the perceived time needed to complete the business model was only marginally higher while the perceived frustration level was significantly lower. In long-term empirical studies it should be tested, whether the derived questions serve their purpose. Furthermore, the taxonomy needs to be validated and reevaluated, since the matching of individual aspects to categories could not be deduced from the literature analysis.

Available business modelling tools omit background domain knowledge, neglecting the users' need for a guided model configuration. Hence, important interdependencies between business model elements indicating weaknesses and threats are not revealed. This jeopardizes the exploitation of an organization's total strategic potential.

As can be derived from the average TLX scores, the developed tool performs significantly better than a non-question-based tool. However, the evaluated criteria can be improved even further. The user interface can be adapted in order to guide the user through the configurator more time-efficiently – ultimately lowering the users' frustration even further.

The underlying **taxonomy** was derived under the premise of being flexible. Thus, the aspects added by the user under "other" will be considered in the taxonomy when crossing a certain threshold in relation to the total number of configured business models. Eventually, these aspects will be available for other users as well. Another possible future step could be to reconstruct the questions based on the gained knowledge through the **users' input**. When using the answers to expand the taxonomy's hierarchy levels, a more detailed representation emerges. Also taking the user profiles into account, the provided answer options for selected questions could be prefiltered depending on the answer of previous questions. When a sufficient number of business models is available in the tool, it will be possible to enable a graph-based analysis. By calculating a distance measure between business models and the companies' user profiles, their similarity can be determined. Hence, considering the companies' financials provided through the user profile, an anonymous **recommendation for suitable aspects** can be given – provided the compared organization is more successful than the active user's one.

7 Acknowledgements

The research in this paper has been funded by the German Federal Ministry of Economic Affairs and Energy (BMWi) within the Mittelstand Digital network in the Mittelstand 4.0 Kompetenzzentrum Kaiserslautern.

References

1. Burmeister, C., Luettgens, D., Piller, F.T.: Business Model Innovation for Industrie 4.0: Why the “Industrial Internet” Mandates a New Perspective. SSRN Electron. J. 0, 1–31 (2015)
2. Chesbrough, H.: Business model innovation: Opportunities and barriers. Long Range Plann. 43, 354–363 (2010)
3. Peppard, J., Ward, J.: The Strategic Management of Information Systems, Building a Digital Strategy. Wiley (2016)
4. Al-Debei, M.M., El-Haddadeh, R., Avison, D.: Defining the Business Model in the New World of Digital Business. Proc AMICS 2008. 1–11 (2008)
5. Osterwalder, A., Pigneur, Y.: Business Model Generation. John Wiley & Sons, New York (2010)
6. Alt, R., Zimmermann, H.-D.: Preface : Introduction to Special Section – Business Models. Electron. Mark. 11, 3–9 (2001)
7. Osterwalder, A., Pigneur, Y.: Clarifying Business Models : Origins , Present , and Future of the Concept. Commun. Assoc. Inf. Syst. 15, 1–125 (2005)
8. Osterwalder, A.: Business model ontology. (2004)
9. Chesbrough, H., Rosenbloom, R.: The role of the business model in capturing value from innovation. Ind. Corp. Chang. 11, 529–555 (2002)
10. Nonaka, I., Takeuchi, H., Umemoto, K.: A theory of organizational knowledge creation. Int. J. Technol. Manag. 11, 25–27 (1996)
11. Canvanizer, https://canvanizer.com
12. Start-Green, https://start-green.net
13. Schuster, K.: E-Consulting: Chancen und Risiken. Oldenbourg Wissenschaftsverlag (2005)
14. Schuman, H., Presser, S.: Questions and Answers in attitude surveys: Experiments on question form, wording and context. Sage (1996)
15. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems. Des. Sci. IS Res. MIS Q. 28, 75–105 (2004)
16. Wilde, T., Hess, T.: Methodenspektrum der Wirtschaftsinformatik: Überblick und Portfoliobildung. , München (2006)
17. Gregor, S., Hevner, A.R.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Q. 37, 337–355 (2013)
18. Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kun-, D., Leimeister, J.M., Loos, P., Spann, M.: Business Models – An Information Systems Research Agenda. Bus. Inf. Syst. Eng. 1–15 (2014)
19. Dorst, W., Heyer, T.: Industrie 4.0 – Deutschland als Vorreiter der digitalisierten Vernetzung von Produkten und Produktionsprozessen. (2014)
20. Burkhart, T., Werth, D., Loos, P.: Analyzing the Business Model Concept – A Comprehensive Classification of Literature. (2011)
21. Magretta, J.: Why Business Models Matter. Harv. Bus. Rev. 3–8 (2002)

22. Albeck, W.: Geschäftsmodellinnovationen für das mittlere Marktsegment. (2016)
23. Baden-Fuller, C., Haefliger, S.: Business Models and Technological Innovation. Long Range Plann. 46, 419–426 (2013)
24. Ematinger, R.: Von der Industrie 4.0 zum Geschäftsmodell 4.0. Springer Fachmedien Wiesbaden, Wiesbaden (2018)
25. Schoormann, T., Behrens, D., Knackstedt, R.: Softwaregestützte Modellierung von Geschäftsmodellen. Lect. Notes Informatics (LNI), Gesellschaft für Inform. 1333–1347 (2016)
26. Houtkoop-Steenstra, H., Houtkoop-Steenstra, J.P.: Interaction and the standardized survey interview. Cambridge University Press (2000)
27. Gault, R.H.: A history of the questionnaire method of research in psychology. Pedagog. Semin. 14, 366–383 (1907)
28. Martin, E., Polivka, A.E.: Diagnostics for redesigning survey questionnaires: Measuring work in the Current Population Survey. Public Opin. Q. 59, 547–567 (1995)
29. Sheatsley, P.B.: Questionnaire construction and item writing. Handb. Surv. Res. 4, 195–230 (1983)
30. Foss, N.J., Saebi, T.: Fifteen Years of Research on Business Model Innovation: How Far Have We Come, and Where Should We Go? J. Manage. XX, 1–28 (2016)
31. Sapsford, R.: Survey Research. Sage Publications (2007)
32. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 50, 904–908 (2006)

The Influence of Situational Factors and Gamification on Intrinsic Motivation and Learning

Vildan Salikutluk[1], Henrik Kampling[1], Bjoern Niehaves[1]

[1] University Siegen, Chair of Information Systems, Siegen, Germany
{vildan.salikutluk}@student.uni-siegen.de
{henrik.kampling,bjoern.niehaves}@uni-siegen.de

Abstract. Immersive virtual reality (iVR) is becoming increasingly popular for learning. But how such learning applications are designed is crucial and determines their success. Designing suitable feedback mechanisms in a learning environment manifests through gamification elements. Nevertheless, previous research has shown that the effect of gamification is ambiguous and depends on several aspects. The setting in which the gamification is used can affect the learner's perception of the feedback and, in turn, their motivation. Since learning systems are usually aimed at increasing the user's learning performance but also their inherent enjoyment of learning, investigating effects on the user's intrinsic motivation is essential. This study proposes a research model, and an experimental approach is outlined in order to examine how situational factors influence the effect of gamification on intrinsic motivation and learning performance in iVR learning environments.

Keywords: Learning, Immersive Virtual Reality, Intrinsic Motivation, Gamification, Situational Factors.

1 Introduction

Virtual Reality applications are becoming increasingly popular since the rise of Oculus Rift (Development Kit 2) in 2016. Especially, immersive Virtual Reality (iVR), realized by presenting a virtual environment through a head-mounted display (HMD) which encloses the user into the virtual world and lets the them interact within it by using hand-held controller, is used in different domains and fields. One area of application for iVR is learning [1]. Due to the nature of this technology, different contexts and scenarios can be displayed allowing the user to be completely involved and immersed into the given setting and task within it. Such applications are already deployed at schools and universities as well as in healthcare and process learning [2]. Not only does immersion and interactivity allow for engagement and fun [3], it also furthers concentration by eliminating external distractions [4]. Such deep involvement with the task presented in the virtual environment shows a positive influence on learning and performance [5].

Another important aspect of learning, especially in iVR, is feedback. Evaluating the user's actions and showing them if and what was performed correctly or could be

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

improved is a key aspect of improving their learning performance and avoiding the habituation of erroneous behavior. A popular approach for realizing such technological feedback within such systems is gamification [6]. Gamification can be found in different settings. By applying game elements to work- or learning-related contexts and tasks it provides feedback in a game-like fashion. It is designed to motivate users by awarding them points, badges and level-ups, etc. for desired, correct behavior within the given context or for the specific task.

The aim of such feedback and reward systems is to achieve the fulfillment of the need for competence. This, in turn, is aimed to increase the intrinsic motivation of the users such that they perform the activity for its own sake [7]. This way, the enjoyment and performance of the user is increased inherently through the performance of the task instead of externally motivated by the expectation of rewards [7]. But how gamification is influencing the user's motivation and performance has rarely been empirically investigated [6]. Whether the intended growth in intrinsic motivation actually takes place by introducing gamification to learning tasks and scenarios is not yet clear [6].

While feedback systems and rewards, as they are used in gamified applications, are usually considered to be extrinsic motivators, it is important that these do not reduce but rather even heighten the intrinsic motivation of users - not only because long-term performance and engagement are dependent on intrinsic motivation but also the user's satisfaction and their overall well-being [8]. Hence, it is essential to investigate how gamification can be utilized such that it does not stifle but foster intrinsic motivation and with that the long-term performance, satisfaction and well-being of the user.

This question is difficult to answer as the effect of gamification varies depending on the context it is used in, how it is designed and realized and what situation it is used in. Mekler et al. demonstrate how situational factors have an impact on the performance of users as well as on their intrinsic motivation [9]: If the feedback in the gamified application is understood and perceived by the user as neutral and merely informational, it positively impacts the user's intrinsic motivation. However, if through the situational factors the feedback is perceived as controlling, it has a negative influence on the user's intrinsic motivation. As a consequence, this shows the importance of not only the feedback itself but also points to the significance of its perception by the user. Investigating how the effect of gamification, especially within an iVR setting, is influenced by situational factors promises to deliver clues as to how the intrinsic motivation of a user as well as their learning performance is affected and how these aspects need to be integrated into an iVR learning setup and application. Therefore, this work-in-progress paper proposes an experimental setup with which the following research question can be addressed:

RQ: How do situational factors influence the effect of gamification on intrinsic motivation and with that learning performance in an iVR learning environment?

To answer this research question, the paper is structured as follows. In Section 2 the research background on iVR, gamification and intrinsic motivation is presented. The hypothesis development and the research model are outlined in section 3, followed by the description of the research method and intended data analysis in section 4. Lastly, in section 5 an initial discussion and an outlook is presented.

2 Background and Literature Research

2.1 Immersive Virtual Reality

Immersive Virtual Reality (iVR) is achieved by using complex interfaces and hardware, such as head-mounted displays (HMD), which allows the user to be completely shut off from their surrounding and exclusively perceive the virtual environment. By using controllers or through hand-gestures the user is able to interact with the virtual world and the objects within it. This kind of immersion can lead to the feeling of (tele)presence in the virtual environment and a sense of actually being in this virtual world [10, 11]. This immersion and sense of presence increases engagement with the virtual environment and in turn with the task within it [5]. Besides this, iVR offers several advantages. Due to that, it is used as the environment of choice for many learning applications. For instance, because of the provided interactivity iVR environments are used as medical training tools for medical students [12]. Another example of such applications are games helping students in certain areas of math like, for example, geometry [13]. The main reasons why iVR is utilized in such cases is that it allows to visualize and display complex and abstract information and explanations in a more tangible and interactive way. Enabling this type of interactivity and good visualization positively influences learning outcomes [5] which is why iVR is a popular tool for realizing learning application and serious games.

2.2 Intrinsic Motivation

Intrinsic Motivation is defined as the drive to do a certain activity for its own sake rather than because of any expected consequences or rewards which follow it [14]. As opposed to extrinsically motivated activity, which is elicited by anticipated external rewards and, therefore, dependent on external factors, intrinsically motivated behavior driven by the inherent joy of the activity. In order to reach and foster this intrinsic type of motivation, it is important to fulfill certain needs. According to the Self-Determination Theory by Ryan and Deci [8], increased intrinsic motivation, and with that, well-being can be achieved by meeting the psychological needs of competence, social relatedness and autonomy. This means that by creating and enabling (social) structures and situations which allow the fulfillment of these needs or at least a part of them, the development of self-motivation and overall healthy mental development and state can be supported [8].

In order to foster intrinsic motivation towards learning, it is important to take these psychological needs into account. Since a crucial part of learning is based on receiving feedback [15], the way in which this feedback is presented can influence the satisfaction and fulfillment of these needs. It is important that external feedback does not diminish but rather support intrinsic motivation. To be able to provide feedback in a helpful fashion such that it can promote feelings of competence and/or autonomy could positively affect the overall motivation to learn. As a consequence, integrating this insight and designing learning applications and feedback mechanisms with this type of need satisfaction in mind can be beneficial.

2.3 Gamification and Situational Factors

Gamification can be understood as the application of game elements to non-game contexts [16]. Which elements are utilized can vary: for example, it is possible to include ranks, badges, levels, etc. [16]. Hamari et al. [3] demonstrated, through their meta-analysis of 24 gamification studies, that such game elements have a positive effect on users. Two effects which were found across all studies are fun and engagement. This is supported by the results from Buckley et al. [17], who also found that gamification positively affects users. However, the impact of each element depends on the context in which it is used as well as how it is implemented [3, 18]. An important game element which is frequently used in gamified applications and which is also highly context-sensitive is feedback [16]. Feedback can be realized and displayed to the user in several forms with the most common and basic type being points [19]. They are the basic building block on which other elements such as level-ups and badges, for instance, are based on. Points are the simplest form of reward and feedback for certain actions within a game or task [6, 20, 21].

How the user perceives this feedback is crucial. If it is understood as purely informational it can support the fulfillment of the psychological needs mentioned above. But if the user experiences it as controlling, it can stifle intrinsic motivation and only be extrinsically motivating [8]. Therefore, if the points and the setting and context in which they are presented to the user are informational, this can lead to an increase in their perceived competence [19, 21]. However, if the situational factors are such that the feedback from the points is considered controlling, they could increase a feeling of pressure and therefore diminish intrinsic motivation.

Mekler et al. [6] showed that the game element of points can not only provide feedback to the user but can increase their intrinsic motivation in an image annotation task. For this, they let the user name and tag images on a computer and provided points for quantity and quality of their tags. Their findings show an increase of intrinsic motivation in subjects who received feedback in the form of points for their actions as opposed to the control group who did not receive any feedback in form of points. However, there are also findings, for instance by Mekler et al. [18] and Sailer et al. [21], which demonstrate that the gamification elements, as points, are not necessarily effective in increasing intrinsic motivation per se.

3 Research Model and Hypothesis Development

In this section the research model is proposed, and the hypotheses will be derived (Figure 1). While previous research does not conclusively show that higher intrinsic motivation leads to increased learning, there are clear tendencies in this direction. Intrinsically motivated subjects often demonstrate at least equal or better learning performances to their extrinsically motivated counterparts. Several studies showed the positive influence of higher intrinsic motivation on learning [22], goal achievement [23] and persistence in education, i.e. lower drop-out rate in education [24]. This points to a positive influence of intrinsic motivation on, especially long-term, learning. Based on this, users of a learning system who exhibit increased intrinsic motivation should be able to show higher learning performance than those who are less intrinsically motivated. This leads to the first hypothesis:

H1: Intrinsic motivation positively influences the learning.

Furthermore, studies like Hamari et al. [3], have clearly demonstrated that the effect of gamification is positive on engagement, fun, and partly on learning outcome [5] and intrinsic motivation [6]. Since the user has to be attentive and is engaged as if they play a game, such positive effects can occur. However, since such results are not conclusive, this needs to be further investigated. In addition, to the best of our knowledge, none of the previously conducted studies on gamification and situational factors are set within and focused on an iVR technologies. Using gamification elements to provide neutral feedback and increase the feeling of competence can lead to fostering intrinsic motivation. Since points are the most basic game element on which previous research builds upon ([19], [21]) it is picked as the element of choice in this study. By combining these aspects, the second hypothesis is:

H2: Providing feedback in an iVR learning environment in form of points, as a gamification element, positively influences intrinsic motivation.

However, the effectiveness of such gamification elements dependent on which game element is used specifically, how it is integrated, and in which context it is embedded in [3, 18, 21]. Therefore, it is important to account for situational factors which could influence the perception and with that the effect of the feedback. Examining situational settings is, therefore, crucial as well. By creating situational settings which can either be considered controlling or informational and neutral allows for an investigation into these situational factors and their impact on the user's intrinsic motivation [8]. Hence, if users perceive the gamification feedback as informational and neutral, they will be more intrinsically motivated than in a controlling setting since no external pressure will affect their performance. This leads to the third hypothesis:

H3a: Adjusting situational factors such that the gamification element is perceived as informational rather than controlling will positively influence intrinsic motivation.

But even if the gamification element and, therefore, a learning or performance feedback is omitted, previous research shows that situational factors still are of important. Situational factors, for example, as the allocation of authority and power, can influence intrinsic interest [25]. In addition, a setting strongly emphasizing the importance of good performance tends to induce pressure and is found to be negatively correlated with intrinsic interest [26]. In contrast, a neutral setting which is focused on learning instead of performance is less pressure-inducing and positively correlated with intrinsic interest [26]. This means, adjusting the situational factors such that they are neutral should lead to higher intrinsic motivation compared to a controlling setting. Therefore, the last hypothesis follows:

H3b: Adjusting situational factors such that they create an informational and neutral setting positively influences intrinsic motivation as opposed to a controlling setting.

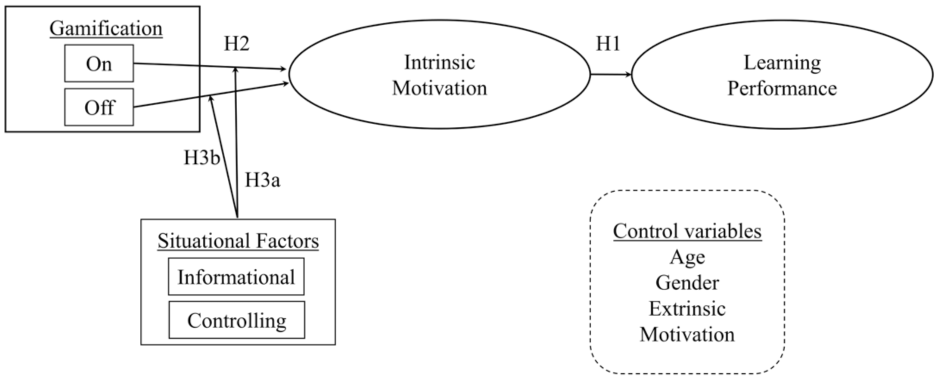


Figure 1. Research Model

4 Method

4.1 Design and Participants

In order to investigate the research question at hand, a randomized 2 (gamification: on vs. off) x 2 (situational factors: informational/neutral vs. controlling) between-subject experimental set up will be used. The experiment will take place in a laboratory in which the participants will be set in a virtual environment, implemented in Unreal Engine 4. For this, an HTC Vive VR-HMD and its controllers will be used. This allows full visual immersion into the virtual environment and interactivity by moving around and move virtual objects. The aim is to recruit 120 participants, with 30 participants per group, with most of them being university students who will receive monetary compensation for their participation in the experiment.

4.2 Learning Task

The learning task used in the experiment is set in a virtual post office. The task is based on the processes in real post offices. This post office is set up in the following way. The participants will stand in front of a desk behind a counter. The customers enter and interact with them straight ahead from this standpoint. In front of them, on the desk, are a scale which allows them to weigh an item and simultaneously show them its measurements like height and length. In addition to that, the stamps and tags for special orders and requests are on the desk. To their left, on the wall, is the price list as well as specific boxes for certain sizes of items. The participants are required to take the role of a post worker interacting with a virtual customer. The customer gives the participant a postcard, letter or package to mail. Which item is brought in by the customer is randomized such that the process varies in each trail. The participant's task is then to execute all necessary steps in order to successfully set up the shipment of the customer's item. That way, the participant is supposed to learn the execution of the customer interaction and shipping process. The steps taught in the experiment at hand will be described in the following:

1. Receive customer item and listen/read their instructions (for example, how should the letter be mailed)
2. If special order: add tag
3. Weigh and measure the customer's item
4. Look up the price of such an item on a price list
5. Bill the customer accordingly by typing in the right price into a number pad
6. Attach stamp to the item
7. Sort the item according to its size and weight into the right shipment box

Occasionally, a customer will have a special order or requirement which adds an extra step (see step 2). At this point, they have to an attach a tag that indicates that this is, for example, express shipment.

The system is designed in a way such that it does not allow for the participant to make a mistake and finish the process regardless of that. Instead, the participant cannot continue with the next steps until the mistake it corrected, and the steps can be followed in the right order again.

4.3 Procedure

The experiment takes place in a virtual reality laboratory and participants are tested individually. Firstly, the participants receive a short introduction by the experimenter regarding the consent form and the general set up of the experiment. If participant decide to take part, they receive oral as well as written instructions by the experimenter regarding the task. After the task is clear and the participant knows what to do, the experimenter explains how the VR Headset and its controllers work and how participants can navigate and interact in the virtual world.

Given that the participant is in the group in which the gamification is tested, they will see a board in the virtual post office on the upper right corner of the virtual desk they are working on. On this board their current point score will be displayed. If the participant is in the non-gamification group of subjects, the board will not be shown. Instead of serving as extrinsic motivators, these points are aimed to increase the subject's perceived competence and in turn intrinsically motivate them.

Depending on which condition the participant is in, they either receive task instructions which are phrased in an informational and neutral or in a controlling way. The instructions are based on the ones used in an experiment by Shalley and Smith [27]. The phrasing is specifically adjusted in each condition with either neutral or pressure-inducing and autonomy-depriving words [28]. In the controlling condition, the participants are also told by the experimenter that they are an important part of the study, that they will be observed for the purpose of evaluating their performance during the task and that the time they take to finish the task will be measured as well. They are explicitly told to make no mistakes while working through the task as fast as they can. By introducing these instructions, we assume that the feeling of pressure will be increased in those subjects. The participants in the other group with neutral instructions are just told that they are a part of the study, that they should just simply try to do their best but that they are in no hurry and if something goes wrong it is not a problem. They also get told that the experimenter will be watching what they are doing, however, they are assured that this is just for their safety so that the experimenter can make sure that they do not bump into objects.

After an initial training period to get used to the environment and to the setting and interactions, the participants are required to start the task and their time and number of errors will be taken by the experimenter. After ten trails are completed, they can stop. Lastly, they get asked to complete a questionnaire for which they have to indicate their degree of agreement with certain statements. These statements are the items which are aimed at measuring the participants level of intrinsic motivation and perceived pressure. The measures are explained in more detail in the section below and the items used in the questionnaire can be found in Table 1 in the Appendix.

4.4 Measures

Learning Performance: To measure the learning outcome or the performance we use the point system in the virtual environment as well as an error count through observations by the experimenter. The system rewards points such that it counts three points for a process completed without mistakes, two points for a process in which one error was committed and one point for every other completed process. In addition, the experimenter will count the number of mistakes by observing the actions in the virtual environment.

Intrinsic Motivation: To measure intrinsic motivation we use the Intrinsic Motivation Inventory (IMI) [29]. Some items are adapted such that they fit this specific experiment (Appendix, Table 1).

Pressure: To measure pressure, the construct of "Pressure" within the Intrinsic Motivation Inventory was used. The items within this construct are also adapted to fit this specific experiment (Appendix, Table 1).

Perceived Learning: To measure the perceived learning of the participants, items by Goel et al. [30] were adjusted (Appendix, Table 1).

Learning Satisfaction: To measure the learning satisfaction of the participants, items by Goel et al. [30] were adjusted (Appendix, Table 1).

Manipulation Checks: Three manipulation checks are introduced. Since the display of the gamification element is manipulated in this experiment, the points are either shown or not presented to the participant. Due to this, the participants are asked whether they could collect and accumulate points from the system for their actions. Additionally, by using adjusted items by Suh et al. [31] it is checked whether the collection of points yielded any motivational incentive for the participants. Lastly, the situational factors are adjusted such that they are either controlling or neutral. Therefore, participants are asked whether the instructions were perceived as either pressure-inducing or neutral. The manipulation checks can be found in Table 1.

4.5 Data Analysis

In order to test the hypotheses and answer the research question different statistical methods will be utilized. For instance, a multiple regression analysis and ANCOVA to check manipulations and whether the participants in each condition (non-/gamification and neutral/controlling situational factors) performed differently and exhibited (significant) distinctions in their level of intrinsic motivation.

5 Discussion

This study is aimed at investigating the influence of situational factors on gamification in an iVR learning environment and their impact on intrinsic motivation and learning outcome. This work will contribute to the body of empirical research revolving around the subjects learning in iVR (for example, [1]), intrinsic motivation (for instance, [6], [8] and [9]), and the influence of situational factors (for example [26]). In addition to that, it will inform and provide insights for practitioners and designers of learning system as well as their users. Arranging and adjusting situational factors could help to motivate the learners and improve their performance. Therefore, questions of how to deploy, use, and get the most out of such systems needs to be answered, whereby the study at hand will contribute by its findings. In order to proceed, the next step is participant recruitment and conducting the experiment.

As every study, this research has certain limitations. Although gamification elements might affect the intrinsic motivation of a person, there might be further design-related and non-design aspect that affect the individuals' motivation. Hence, future research should extend this perspective by including other factors as well. For instance, future research might investigate alternative gamification elements (e.g., batches, leaderboard, etc.) as a promising aspect to influence learning performance and variation in motivation in more detail. Finally, since an experimental approach is proposed here, external validity might be an issue which could be considered in further studies.

6 Acknowledgements

This article was supported by project ELISE and by a grant provided by the German Federal Ministry of Education and Research (BMBF) (No. 16SV7512).

References

1. Martín-Gutiérrez, J.: Virtual Technologies Trends in Education. EURASIA Journal of Mathematics, Science and Technology Education. 13, (2016).
2. Hayden, S.: Volkswagen Group to Train 10,000 Employees in VR This Year. *Road to VR*. (2018).
3. Hamari, J., Koivisto, J., Sarsa, H.: Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. 2014 47th Hawaii International Conference on System Sciences. (2014).
4. Agarwal, R., Karahanna, E.: Time Flies When Youre Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly. 24, 665 (2000).
5. Dede, C.: Immersive Interfaces for Engagement and Learning. Science. 323, 66–69 (2009).
6. Mekler, E.D., Brühlmann, F., Opwis, K., Tuch, A.N.: Disassembling gamification. CHI 13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA 13. (2013).
7. Csikszentmihalyi, M.: Play and Intrinsic Rewards. Flow and the Foundations of Positive Psychology. 135–153 (2014).
8. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist. 55, 68–78 (2000).
9. Mekler, E.D., Brühlmann, F., Opwis, K., Tuch, A.N.: Do points, levels and leaderboards harm intrinsic motivation? Proceedings of the First International Conference on Gameful Design, Research, and Applications - Gamification 13. (2013).
10. Schultze, U.: Embodiment and presence in virtual worlds: a review. Journal of Information Technology. 25, 434–449 (2010).
11. Schultze, U.: Performing embodied identity in virtual worlds. European Journal of Information Systems. 23, 84–95 (2014).
12. Müns, A., Meixensberger, J., Lindner, D.: Evaluation of a novel phantom-based neurosurgical training system. Surgical Neurology International. 5, 173 (2014).
13. Hwang, W.-Y., Hu, S.-S.: Analysis of peer learning behaviors using multiple representations in virtual reality and their impacts on geometry problem solving. Computers & Education. 62, 308–319 (2013).
14. Ryan, R.M., Deci, E.L.: Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology. 25, 54–67 (2000).
15. Hatziapostolou, T., Paraskakis, I.: Enhancing the impact of formative feedback on student learning through an online feedback system. Electronic Journal of e-Learning. 8(2), 111-122 (2010).
16. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness. Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek 11. (2011).
17. Buckley, P., Doyle, E.: Gamification and student motivation. Interactive Learning Environments. 24, 1162–1175 (2014).

18. Mekler, E.D., Brühlmann, F., Tuch, A.N., Opwis, K.: Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior. 71, 525–534 (2017).
19. Aparicio, A.F., Vela, F.L.G., Sánchez, J.L.G., Montes, J.L.I.: Analysis and application of gamification. Proceedings of the 13th International Conference on Interacción Persona-Ordenador - INTERACCION 12. (2012).
20. Sailer, M., Hense, J., Mandl, H., Klevers, M.: Psychological perspectives on motivation through gamification. IxD&A, 19, 28-37 (2013).
21. Sailer, M., Hense, J.U., Mayr, S.K., Mandl, H.: How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior. 69, 371–380 (2017).
22. Cordova, D.I., Lepper, M.R.: Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. Journal of Educational Psychology. 88, 715–730 (1996).
23. Curry, S., Wagner, E.H., Grothaus, L.C.: Intrinsic and Extrinsic Motivation for Smoking Cessation. PsycEXTRA Dataset. (1990).
24. Vallerand, R.J., Fortier, M.S., Guay, F.: Self-determination and persistence in a real-life setting: Toward a motivational model of high school dropout. Journal of Personality and Social Psychology. 72, 1161–1176 (1997).
25. Ames, C.: Achievement goals, motivational climate and motivational processes. In: Roberts, G.C. (ed.) Motivation in Sport and Exercise. pp. 161–176.
26. Cury, F., Biddle, S., Famose, J.P., Sarrazin, P., Durand, M., Goudas, M.: Personal and Situational Factors Influencing Intrinsic Interest of Adolescent Girls in School Physical Education: a structural equation modelling analysis. Educational Psychology. 16, 305–315 (1996).
27. Shalley, C.E., Perry-Smith, J.E.: Effects of Social-Psychological Factors on Creative Performance: The Role of Informational and Controlling Expected Evaluation and Modeling Experience. Organizational Behavior and Human Decision Processes. 84, 1–22 (2001).
28. Radel, R., Pelletier, L.G., Sarrazin, P., Milyavskaya, M.: Restoration process of the need for autonomy: The early alarm stage. Journal of Personality and Social Psychology. 101, 919–934 (2011).
29. Markland, D., Hardy, L.: On the Factorial and Construct Validity of the Intrinsic Motivation Inventory: Conceptual and Operational Concerns. Research Quarterly for Exercise and Sport. 68, 20–32 (1997).
30. Goel, L., Johnson, N.A., Junglas, I., Ives, B.: How cues of what can be done in a virtual world influence learning: An affordance perspective. Information & Management. 50, 197–206 (2013).
31. Suh, A., Cheung, C.M., Ahuja, M., Wagner, C.: Gamification in the Workplace: The Central Role of the Aesthetic Experience. Journal of Management Information Systems. 34, 268–305 (2017)

Appendix

Table 1. Constructs and Examples of Items for Experiment.
For all items, participants have to indicate to which degree they agree with the given statement on a scale from 1 (not agree at all) to 7 (completely agree).

Construct		**Items - Examples**
Intrinsic Motivation Inventory (IMI) [29]	Enjoyment/Interest	I thought the task was quite enjoyable. I would describe the task as very interesting.
	Perceived Competence	I am satisfied with my performance at this task. I think I am pretty good at the task.
	Perceived Choice	I believe I had some choice about doing the task. I did the task because I wanted to.
	Effort/Importance	I put a lot of effort into doing the task. It was important to me to do well at this task.
	Pressure/Tension	I felt very tense while doing this task. I felt pressured while doing the task.
Perceived Learning [30]		Doing the task increased my knowledge about the process.
Learning Satisfaction [30]		I am satisfied with the way I learned about the process.
Awareness of Point System (MC) [self-developed]		The system offers me the possibility to accumulate points I have gained.
Points and Motivation (Manipulation Check) [self-developed]		The system motivated me by offering the possibility to obtain more points if I try hard.
Situational Factors (Manipulation Check) [self-developed]		I felt pressured by how I was instructed to do the task.

Evaluation von ITSM-Tools für Integration und Management von Cloud-Diensten am Beispiel von ServiceNow

Dominik Schneider[1,2], Franziska Plate[1,2] und Gunnar Auth[2]

[1] Detecon International GmbH, IT Management, Köln, Deutschland
{dominik.schneider,franziska.plate}@detecon.com
[2] Hochschule für Telekommunikation Leipzig, Institut für Wirtschaftsinformatik, Leipzig, Deutschland
auth@hft-leipzig.de

Abstract. Im Kontext der digitalen Transformation nimmt die Nutzung von IT-Services aus der Cloud in Unternehmen rasant zu. Diese Entwicklung führt zu neuen Herausforderungen im IT Service Management (ITSM). Das auf das Management interner IT-Services ausgerichtete ITSM muss erweitert werden, um externe IT-Services aus der Cloud zu integrieren und zu managen. In Literatur und Praxis wird Service Integration and Management (SIAM) als vielversprechender Ansatz zur Erweiterung des ITSM um diese Fähigkeit diskutiert. Dabei stellt sich auch die Frage, wie sich geeignete ITSM-Tools ermitteln lassen, die Prozesse und Komponenten von SIAM mit ihrem Funktionsumfang abdecken. Diese Fragestellung wird mittels kriterienbasierter Evaluation am Beispiel eines ITSM-Tools des Herstellers ServiceNow untersucht. Basierend auf den Ergebnissen der Evaluation werden Aussagen über die allgemeine Abdeckung von ITSM-Tools hinsichtlich der Umsetzung von SIAM abgeleitet und diskutiert.

Keywords: IT Service Management, Service Integration and Management, SIAM, Evaluation, ITSM-Tools

1 Einleitung

Im Kontext der digitalen Transformation, womit fundamentale Veränderungen unter anderem der Geschäftswelt durch disruptive IT-Innovationen gemeint sind [1], nimmt die Nutzung von IT-Services aus der Cloud in Unternehmen rasant zu. Laut Bitkom nutzten im Jahr 2016 bereits 65 Prozent der deutschen Unternehmen Software, Speicher oder Rechenleistung aus der Cloud [2]. Trotz der stark gestiegenen Verbreitung ruft das Cloud-Computing bei IT-Entscheidern auch kritische Reaktionen hervor [3]. Ein Grund dafür ist, dass die Nutzung von IT-Services aus der Cloud in Unternehmen zu weitreichenden Veränderungen im IT Service Management (ITSM) führt. Vor diesem Hintergrund hat sich unter der Bezeichnung Service Integration and Management (SIAM) ein neuerer ITSM-Ansatz herausgebildet, der dazu dient, IT-Services aus der Cloud zu integrieren und ganzheitlich zu steuern [4]. SIAM verfolgt damit das

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Ziel, eine einheitliche geschäftsorientierte IT-Organisation abzubilden. Das Konzept kann als Erweiterung des herkömmlichen ITSM nach IT Infrastructure Library (ITIL) um spezifische Aspekte des Multiprovider-Managements verstanden werden [4].

Damit SIAM in der ITSM-Praxis einen zusätzlichen Nutzen entfalten kann, müssen ITSM-Tools erweiterte Anforderungen erfüllen, die sich aus den bei SIAM beschriebenen Fähigkeiten ergeben. Unternehmen können am Markt aus einer Vielzahl unterschiedlicher ITSM-Tools auswählen [5]. Obwohl die Erweiterung des ITSM um die Fähigkeiten von SIAM bereits seit etwa 2005 [6] diskutiert wird, scheint sich die Umsetzung von SIAM in die Praxis noch in einem Frühstadium mit zögerlichem Fortschritt zu befinden. Unter der Annahme, dass der Abdeckungsgrad von SIAM-Funktionalitäten in am Markt erhältlichen ITSM-Tools dabei eine wichtige Rolle spielt, wird in der vorliegenden Arbeit eine kriterienbasierte Evaluation am Beispiel des ITSM-Tools Kingston [7] des Herstellers ServiceNow entwickelt und durchgeführt.

Die Arbeit ist wie folgt strukturiert: Nach der Einleitung werden in Abschnitt 2 verwandte Forschungsarbeiten zu SIAM und ITSM diskutiert. Aus einer semiformalen SIAM-Beschreibung [8] wird in Abschnitt 3 der Gesamtansatz in einzelne Komponenten untergliedert, die für die Durchführung der Evaluation relevant sind und als Basis für einen Kriterienkatalog dienen. Anschließend wird die Evaluation am Beispiel des ITSM-Tools Kingston von ServiceNow exemplarisch durchgeführt (Abschnitt 4) und diskutiert (Abschnitt 5). Die Arbeit schließt mit einem Fazit und Ausblick in Abschnitt 6.

2 Stand der Forschung

Es existieren verschiedene Untersuchungen, die sich mit ITSM im Umfeld des Cloud-Computings befassen. Pröhl et al. [3] setzen sich mit den Prozessveränderungen im ITSM auseinander, die sich aus der zunehmenden Nutzung von IT-Services aus der Cloud in Unternehmen ergeben. Dabei wird die Bedeutung dieser durch das Cloud-Computing veränderten bzw. zu verändernden Prozesse am Beispiel von ITIL aufgezeigt. Abschließend wird eine Neubewertung der ITSM-Prozesse in Cloud-Computing-Szenarien vorgenommen. Heininger et al. [9] führen eine Literaturrecherche zu ITSM im Umfeld des Cloud-Computings durch. Sie zeigen, dass in der untersuchten Literatur die Notwendigkeit einer Erweiterung des ITSM zur Integration und zum Management von IT-Services aus der Cloud thematisiert wird. Zudem wird betont, dass im ITSM Handlungsbedarf besteht, um einheitliche geschäftsorientierte IT-Organisationen bestehend aus internen und externen IT-Services zu realisieren. Wie dieser Handlungsbedarf allerdings konkret aussieht, wird nicht erläutert. Robrecht et al. [10], Liebhart und Wischki [11] sowie Arabalidousti et al. [12] unterstreichen, dass in der Literatur zum Großteil nur theoretisch auf ITSM im Zusammenhang mit Cloud-Computing eingegangen wird. In diesem Kontext werden häufig identische oder ähnliche Handlungsbedarfe aufgezeigt und begründet, die praktische Auseinandersetzung mit diesen Handlungsbedarfen jedoch als weitere Forschungsthemen in Ausblick gestellt.

Neuere Quellen beschäftigen sich mit SIAM, um verstärkt die Handlungsbedarfe der ITSM-Praxis im Multiprovider-Umfeld anzugehen. Von Holland [8, 13] werden Prozesse beschrieben, die aus spezifischen Komponenten bestehen und die Integration sowie das Management von IT-Services ermöglichen. Nägele et al. [14] identifizieren kritische Erfolgsfaktoren des IT-Multisourcings mit SIAM. Hallikainen [15] befasst sich mit der Überlegung, einen Service-Katalog und die Configuration Management Data Base (CMDB)-Struktur als Basis für die Verknüpfung von ITSM und SIAM zu nutzen. Ein konkretes Vorgehen für die Einführung und Anwendung von SIAM in Unternehmen beschreiben Armes et al. [16] und Virri [17]. Armes et al. gehen auf Grundsätze und Vorgehensweisen für die Einführung von SIAM ein. Darauf aufbauend fokussiert sich Virri [17] auf den Einsatz der Vorgehensweisen in konkreten Anwendungsfällen. Lehtonen [18] thematisiert praktische Leitlinien zur Auswahl eines geeigneten SIAM-Modells.

Einige dieser Beiträge betrachten SIAM weitgehend isoliert von herkömmlichen ITSM-Ansätzen. Es ist jedoch sinnvoll, SIAM und ITSM gemeinsam zu untersuchen, da die SIAM-Prozesse und -Komponenten erst dann einen Mehrwert bieten, wenn sie zur Beherrschung des Multiprovider-Managements in das bestehende ITSM integriert werden. Im Umgang mit IT-Services aus der Cloud gewinnt das Multiprovider-Management zunehmend an Bedeutung. Vor diesem Hintergrund erwächst aus der komplementären Anwendung von SIAM und ITSM ein großes Nutzenpotenzial.

In der Literatur werden verschiedene Möglichkeiten betrachtet, um SIAM auf Toolebene in das ITSM zu integrieren. Eine Möglichkeit ist die Entwicklung eigenständiger Tools basierend auf SIAM. Hintergrund dieser Überlegung ist es, Inkompatibilitäten zwischen den von IT-Service-Anbietern und IT-Service-Empfängern genutzten ITSM-Tools zu überwinden, um heterogene IT-Services zu integrieren und zu managen [19]. Zur Verknüpfung unterschiedlicher ITSM-Tools bei IT-Service-Anbietern und IT-Service-Empfängern lassen sich dedizierte SIAM-Tools einsetzen, wie bspw. ServiceFlow [19]. Eine andere Möglichkeit ist es, ausgewählte SIAM-Komponenten in klassischen ITSM-Tools zu implementieren. Es gibt bereits ITSM-Tool-Anbieter, die ihre ITSM-Tools um SIAM-Komponenten erweitert haben [19].

3 Konzeption der kriterienbasierten Evaluation

Die Beurteilung informationstechnischer Artefakte wie bspw. Software gilt als ein spezifisches Evaluationsproblem der Wirtschaftsinformatik [20]. Heinrich [21] begründet die Notwendigkeit eines einheitlichen Evaluationsverfahrens und gibt Hinweise für die Ableitung von Kriterien aus dem Evaluationsobjekt. Das im folgenden beschriebene Verfahren basiert zudem auf dem bei Krcmar [22] beschriebenen allgemeinen Verfahren zur Softwareauswahl. Evaluationsziel ist die Bewertung von kommerziellen ITSM-Tools hinsichtlich ihrer Eignung für die Unterstützung einer IT-Organisation gemäß SIAM. Dazu werden aus einer semiformalen SIAM-Beschreibung [8] einzelne SIAM-Komponenten abgeleitet und in einen Kriterienkatalog überführt.

Die Konstruktion des Kriterienkatalogs erfolgt in zwei Schritten. Im ersten Schritt werden aus der SIAM-Beschreibung [8] drei Domänen abgeleitet, die die horizontale Struktur des Katalogs festlegen. Domänen definieren einen thematischen Zusammenhang für die darin zusammengefassten Komponenten. Im zweiten Schritt werden die Domänen durch SIAM-Prozesse und dazugehörige SIAM-Komponenten weiter strukturiert. Prozesse unterteilen Domänen in verschiedene Themenbereiche und gruppieren Komponenten. Durch die weitere Strukturierung der Domänen entsteht eine vertikale Struktur bestehend aus drei Detailstufen. Domänen beschreiben die geringste Detaillierung, Prozesse stellen die zweite Detailstufe dar. Die SIAM-Komponenten bilden die feinste Detailstufe des Katalogs und werden als Vergleichskriterien für die Durchführung der Evaluation verwendet.

Abbildung 1 zeigt die aus der SIAM-Beschreibung abgeleiteten Domänen und Prozesse. Domäne 1 beschreibt übergreifende Prozesse, die sich schwerpunktmäßig mit der Geschäfts- und Kundenorganisation, sowie dem Beziehungsmanagement befassen. Sie sind ebenfalls für die anderen beiden Domänen relevant. Domäne 2 fokussiert sich auf Prozesse, welche zur Integration und zum Management extern bereitgestellter IT-Services notwendig sind. Domäne 3 legt den Schwerpunkt auf das Management interner IT-Services.

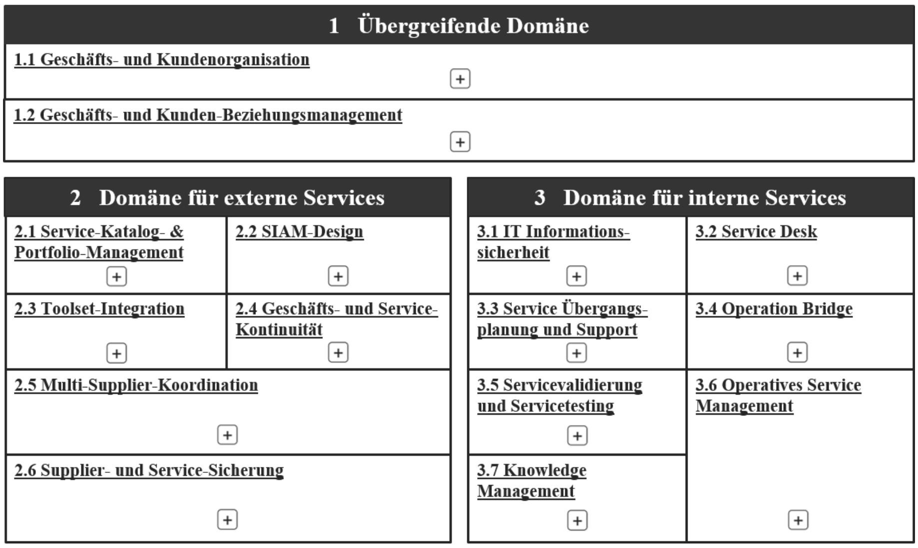


Abbildung 1. Domänen und Prozesse des Kriterienkatalogs (in Anlehnung an [8])

Da in der vorliegenden Arbeit untersucht wird, inwieweit die Funktionalitäten von ITSM-Tools SIAM zur Integration und zum Management von IT-Services aus der Cloud abdecken, liegt der Fokus auf extern bereitgestellten IT-Services. Aus diesem Grund wird im Folgenden die zweite Domäne fokussiert, weiter detailliert und die entsprechenden SIAM-Komponenten werden zur Durchführung der Evaluation verwendet. Die Eingrenzung des Fokus der Evaluation auf Domäne 2 erfolgt außerdem, um das Evaluationsvorgehen anhand eines konkreten ITSM-Tools zu evaluieren und die Qualität der Analyseergebnisse zu beurteilen. Die Erweiterung des Scopes auf die Domänen 1 und 3 ist denkbar, jedoch nicht Gegenstand dieser Arbeit.

Die Detailstufe des in Abbildung 1 abgebildeten Kriterienkatalogs ist für die Durchführung einer Evaluation nicht ausreichend. Es werden detailliertere Zielvorgaben der Domäne für extern bereitgestellte IT-Services benötigt. Der Detaillierungsgrad der Zielvorgaben muss ähnlich sein wie der der Funktionalitäten des in der Evaluation zu analysierenden ITSM-Tools. Deshalb werden die Prozesse 2.1 bis 2.6 weiter zerlegt und relevante dazugehörige SIAM-Komponenten aus [8] zugeordnet. Die ausgewählten Komponenten bilden die dritte Detailstufe des Katalogs ab. Die Prozesse 2.1 bis 2.6 und deren Komponenten sind in Abbildung 2 aufgeführt.

2 Domäne für externe Services	
2.1 Service-Katalog- & Portfolio-Management 2.1.1 Service-Katalog-Management 2.1.2 Service-Portfolio-Management	**2.2 SIAM Design** 2.2.1 Design-Abstimmung 2.2.2 Design des Verfügbarkeitsmanagements 2.2.3 Design des Kapazitätsmanagements
2.3 Toolset-Integration 2.3.1 Unterstützung der Prozessausführung 2.3.2 Service-Warnungen und -Monitoring 2.3.3 Diagnosen 2.3.4 Sicherheit 2.3.5 Reporting 2.3.6 Analysen 2.3.7 Entscheidungsunterstützung	**2.4 Geschäfts- und Service-Kontinuität** 2.4.1 IT-Service Kontinuitätsmanagement 2.4.2 Geschäftskontinuitätsmanagement
2.5 Multi-Supplier-Koordination 2.5.1 Change Management 2.5.2 Release-Planung und -Konfliktmanagement 2.5.3 Kapazitätsmanagement 2.5.4 Major-Incident-Management 2.5.5 Problemmanagement 2.5.6 Innovation 2.5.7 Kontinuierliche Serviceverbesserung	
2.6 Supplier- und Service-Sicherung 2.6.1 Service Level Management 2.6.2 Audits 2.6.3 Bewertung von Prozessreife und Fähigkeiten 2.6.4 Bewertung von Supplierreife und Fähigkeiten 2.6.5 Test Sicherung des Supplier-Testings 2.6.6 Compliance-Audits anhand der Anforderungen 2.6.7 Compliance-Audits anhand von Standards 2.6.8 Monitoring von kontinuierlichen Serviceverbesserungs-Initiativen	

Abbildung 2. Prozesse und Komponenten der Domäne für extern bereitgestellte Services (in Anlehnung an [8])

Die insgesamt 29 Komponenten der Prozesse 2.1 bis 2.6 bilden zusammen die Prüfkriterien, anhand derer untersucht wird, inwieweit ein ITSM-Tool SIAM zur Integration und zum Management von IT-Services aus der Cloud bereits abdeckt. Die Komponenten definieren die Sollvorgabe der Evaluation. Jede einzelne Komponente muss vom ausgewählten ITSM-Tool implementiert werden, damit von einer vollständigen Abdeckung in Bezug auf die Ausgangsfrage gesprochen werden kann. Neben der Sollvorgabe für die Evaluation wird ein Ist-Zustand benötigt, der den Sollvorgaben zum Vergleich gegenübergestellt wird. Der Ist-Zustand ist charakterisiert durch Funktionalitäten eines ausgewählten Tools und wird nachfolgend definiert.

Für die Auswahl des ITSM-Tools wurde der sog. Magic Quadrant für ITSM-Tools von Gartner [5] herangezogen, der auch in der Praxis häufig als erste Orientierung für eine Auswahlentscheidung dient. Der Magic Quadrant kategorisiert ITSM-Tool-Anbieter entsprechend der Dimensionen Vision und Umsetzungsfähigkeit in Nischen-Anbieter, Visionäre, Herausforderer und Führer. Dem Magic Quadrant ist zu entnehmen, dass ServiceNow und BMC führende Anbieter von ITSM-Tools sind. Da ServiceNow zudem in den letzten Jahren den Marktanteil bei ITSM-Tools auf Kosten von BMC stark ausbauen konnte, wird das ITSM-Tool Kingston [7] von ServiceNow für die Durchführung des beschriebenen Evaluationsverfahrens ausgewählt.

Kingston gliedert sich in 17 Funktionalitätsbereiche, von denen IT Service Management und Cloud-Management zwei Beispiele sind. Aus den detaillierten Beschreibungen der Funktionalitätsbereiche wurden insgesamt 138 für die Evaluation relevante Funktionalitäten abgeleitet [7]. Darunter fallen beispielsweise das Service-Portfolio-Management und Change-Management aus dem Funktionalitätsbereich IT Service Management oder das Servicekatalog-Management und Cloud-Account-Management aus dem Funktionalitätsbereich Cloud-Management. Am Beispiel des Service-Katalog-Managements wird deutlich, dass es sich bei einigen Funktionalitäten um klassische ITSM-Funktionalitäten handelt, die im Rahmen von Kingston auch SIAM-Aufgaben abdecken. Anhand der Funktionalitätsbeschreibungen ist jedoch nicht für alle Funktionalitäten erkennbar, ob sie lediglich klassische ITSM-Aufgaben oder auch SIAM-Aufgaben abdecken, oder umgekehrt. Es wird daher die Annahme getroffen, dass klassische ITSM-Funktionalitäten und einige andere Funktionalitäten übergreifend sind und auch für andere Funktionalitätsbereiche verwendet werden können.

Die 138 abgeleiteten Funktionalitäten, die eine ähnliche Detailstufe wie die SIAM-Komponenten im Kriterienkatalog aufweisen, stellen den für die Evaluation benötigten Ist-Zustand dar und schließen die Vorbereitung der Evaluation ab.[1]

4 Exemplarische Durchführung der kriterienbasierten Evaluation

Die Durchführung der Evaluation besteht aus drei Schritten. Im ersten Schritt wird eine Tabelle erstellt, die als Arbeitshilfe für die Evaluation fungiert. Im zweiten Schritt wird die Abdeckung der SIAM-Komponenten durch die Kingston-

[1] Der vollständige Kriterienkatalog inkl. der Bewertungen für die untersuchte Kingston-Version steht interessierten Lesern online zur Verfügung: http://www1.hft-leipzig.de/auth/download/2018-11-30_SIAM_Evaluation_ServiceNow.xlsx

Funktionalitäten mithilfe der Tabelle analysiert (Komponentenabdeckung) und auf SIAM-Prozessebene zusammengefasst (Prozessabdeckung). Der dritte Schritt umfasst die Aggregation der einzelnen Prozessabdeckungen und die Berechnung eines kumulierten Abdeckungsgrads für Kingston. Der kumulierte Abdeckungsgrad gibt an, wie viel Prozent der zweiten Domäne des SIAM-Ansatzes Kingston abdeckt. Die untenstehende Abbildung 3 zeigt einen Ausschnitt aus der Tabelle, mit der die Evaluation durchgeführt wurde. Anhand dieses Ausschnitts werden die einzelnen Schritte und Ergebnisse der Evaluation nachfolgend detaillierter erläutert.

SIAM-Komponenten für externe Services (29)

Kingston-Funktionalitäten (138)

	2.1 Service-Katalog- & Portfolio-Management		**...**	**2.6 Supplier- und Service-Sicherung**							
	2.1.1	2.1.2	...	2.6.1	2.6.2	2.6.3	2.6.4	2.6.5	2.6.6	2.6.7	2.6.8
Service-Katalog-Management	●										
Service-Portfolio-Management		●									
Service-Level-Management				●							
Audit-Management					●				●	●	
Lieferanten-Performance-Management						◑	◑				
Workflow-Management						◑					
Test-Management								◑			
Policy- und Compliance-Management									●	●	
Knowledge-Management											◑
Identifikation potentieller Serviceunterbrechungen											◑
Reporting											◑
Troubleshooting-Management											◑
...											
Komponentenabdeckung	●	●	...	●	●	◑	◑	◑	●	●	◑
Prozessabdeckung	**2 / 2 = 100%**		...	**6 / 8 = 75%**							

● Funktionalität deckt Komponente ab ◑ Funktionalität deckt Komponente mit Unsicherheit ab

Abbildung 3. Ausschnitt der Tabelle zur Durchführung der kriterienbasierten Evaluation

Wie Abbildung 3 andeutet, sind in der ersten Spalte der vollständigen Tabelle alle identifizierten 138 Kingston-Funktionalitäten aufgelistet. In der ersten Zeile sind alle Prozesse (2.1 bis 2.6) der Domäne für extern bereitgestellte Services aufgeführt, die gemeinsam alle 29 SIAM-Komponenten gruppieren (angedeutet in der zweiten Zeile). Der abgebildete Ausschnitt stellt nur die Komponenten der Prozesse 2.1 und 2.6, sowie 12 ausgewählte Kingston-Funktionalitäten dar, um den Rahmen für die Evaluation zu erklären. Die letzten beiden Zeilen der Tabelle umfassen die entsprechenden

Komponenten- und Prozessabdeckungen und werden für die Auswertung der Evaluation benötigt.

Bevor die Komponenten- und Prozessabdeckungen ermittelt werden können, muss im zweiten Schritt der Evaluation analysiert werden, inwieweit jede SIAM-Komponente durch die Kingston-Funktionalitäten abgedeckt wird. Um den Grad der Abdeckung einer SIAM-Komponente durch eine Kingston-Funktionalität darzustellen, werden zwei Ausprägungen von Harvey Balls verwendet. Ein voll ausgefüllter Harvey Ball bedeutet, dass eine Kingston-Funktionalität eine SIAM-Komponente abdeckt. Ein halb ausgefüllter Harvey Ball bedeutet, dass eine Kingston-Funktionalität eine SIAM-Komponente mit einer Unsicherheit abdeckt. Diese Variante eines Harvey Balls ist notwendig, da aus den Beschreibungen der Kingston-Funktionalitäten nicht immer eindeutig hervorgeht, ob eine bestimmte SIAM-Komponente abgedeckt wird. Die Verwendung der Harvey Balls sowie die Zuordnung von Kingston-Funktionalitäten zu SIAM-Komponenten werden im Folgenden beispielhaft erklärt.

Das Service-Katalog-Management (2.1.1) ist als SIAM-Komponente definiert, die alle den Nutzern bereitgestellten Services inklusive deren Preise, Verfügbarkeiten, Service Levels, Ressourcenprofile und Einschränkungen beschreibt [8]. ServiceNow [7] beschreibt das Service-Katalog-Management als Funktionalität, die es Nutzern ermöglicht, Servicekataloge zu erstellen, Angebote für Katalogeinträge wie Services und Produkte anzufordern sowie die Qualität und Verfügbarkeit von Services im Katalog sicherzustellen. Da beide Beschreibungen die gleichen übergeordneten Inhalte (Katalogerstellung inklusive Service-Beschreibung, Ressourcenbeschaffung, Qualitäts- und Verfügbarkeitsmanagement) enthalten, wird gefolgert, dass die SIAM-Komponente Service-Katalog-Management von der Kingston-Funktionalität abgedeckt wird. Dies wird in der entsprechenden Zelle durch einen vollen Harvey Ball abgebildet. Die SIAM-Komponente Bewertung von Prozessreife und Fähigkeiten (2.6.3) wird von keiner Kingston-Funktionalität abgedeckt. Es existieren allerdings die beiden Funktionalitäten Lieferanten-Performance-Management und Workflow-Management, die die genannte SIAM-Komponente mit einer Unsicherheit abdecken. Aus der Beschreibung der Funktionalitäten geht nicht eindeutig hervor, mit welchem Vollständigkeitsgrad sie die SIAM-Komponente abdecken. Aus diesem Grund sind in den entsprechenden Zellen halb ausgefüllte Harvey Balls platziert.

Nachdem die Analyse der Abdeckung aller SIAM-Komponenten erfolgt ist, gilt es im dritten Schritt, die Komponenten- und Prozessabdeckungen auszuwerten. Die Komponentenabdeckung fasst für jede Spalte und somit für jede SIAM-Komponente zusammen, ob sie durch die Kingston-Funktionalitäten abgedeckt wird oder nicht. Sofern in einer Spalte mindestens ein voll ausgefüllter Harvey Ball eingetragen ist und die SIAM-Komponente demnach von mindestens einer Kingston-Funktionalität abgedeckt wird, wird in der Zeile der Komponentenabdeckung für die betrachtete SIAM-Komponente ein voll ausgefüllter Harvey Ball eingetragen. Das ist zum Beispiel bei der SIAM-Komponente 2.1.1 der Fall. Wenn in einer Spalte kein voll ausgefüllter, aber mindestens ein halb ausgefüllter Harvey Ball platziert ist, wird für die entsprechende SIAM-Komponente in der Zeile der Komponentenabdeckung ein halb ausgefüllter Harvey Ball eingetragen, wie beispielsweise bei der SIAM-Komponente

2.6.3. Das bedeutet, dass die SIAM-Komponente von den Kingston-Funktionalitäten mit einer Unsicherheit abgedeckt wird. Ist in einer Spalte kein Harvey Ball eingetragen, bedeutet dies, dass die SIAM-Komponente von Kingston nicht abgedeckt wird. In diesem Fall wird in der entsprechenden Zelle kein Harvey Ball eingetragen. Es ist wichtig zu betonen, dass die Komponentenabdeckung nur durch eine der drei beschriebenen Repräsentationen (voller, halber oder kein Harvey Ball) ausgedrückt wird und die Bewertungen in einer Spalte nicht kumuliert werden. Sind in einer Spalte beispielsweise zwei voll ausgefüllte Harvey Balls eingetragen (siehe SIAM-Komponenten 2.6.6 oder 2.6.7), wird die Komponentenabdeckung durch einen und nicht durch zwei voll ausgefüllte Harvey Balls repräsentiert.

Die Prozessabdeckung in der letzten Zeile der Tabelle drückt aus, wieviel Prozent der SIAM-Komponenten eines SIAM-Prozesses durch die Kingston-Funktionalitäten abgedeckt werden. Um eine Aussage darüber zu treffen, werden alle Komponentenabdeckungen eines SIAM-Prozesses kumuliert und durch die Anzahl der SIAM-Komponenten im betrachteten Prozess dividiert. Die Berechnung der Prozessabdeckung ist in Abbildung 3 anhand der SIAM-Prozesse 2.1 und 2.6 verdeutlicht.

Der SIAM-Prozess 2.1 besteht aus zwei SIAM-Komponenten (2.1.1 und 2.1.2), deren Komponentenabdeckung jeweils durch einen voll ausgefüllten Harvey-Ball repräsentiert wird. In diesem Prozess werden folglich zwei von zwei SIAM-Komponenten durch die Kingston-Funktionalitäten abgedeckt, weshalb die Prozessabdeckung 100 Prozent ergibt. Der SIAM-Prozess 2.6 besteht aus acht SIAM-Komponenten (2.6.1 bis 2.6.8). Insgesamt weisen vier SIAM-Komponenten eine volle Komponentenabdeckung (2.6.1, 2.6.2, 2.6.6 und 2.6.7) und vier eine halbe Komponentenabdeckung (2.6.3, 2.6.4, 2.6.5 und 2.6.8) auf. Das bedeutet, dass vier der acht SIAM-Komponenten nur mit einer Unsicherheit durch die Kingston-Funktionalitäten abgedeckt werden. Kumuliert man die vier voll und vier halb ausgefüllten Harvey Balls, ergibt sich eine Summe von sechs Harvey Balls. Zusammengefasst lässt sich folglich die Aussage treffen, dass in dem SIAM-Prozess 2.6 insgesamt sechs von acht möglichen Harvey Balls ausgefüllt sind. Es werden demnach 75 Prozent der SIAM-Komponenten durch die Kingston-Funktionalitäten abgedeckt.

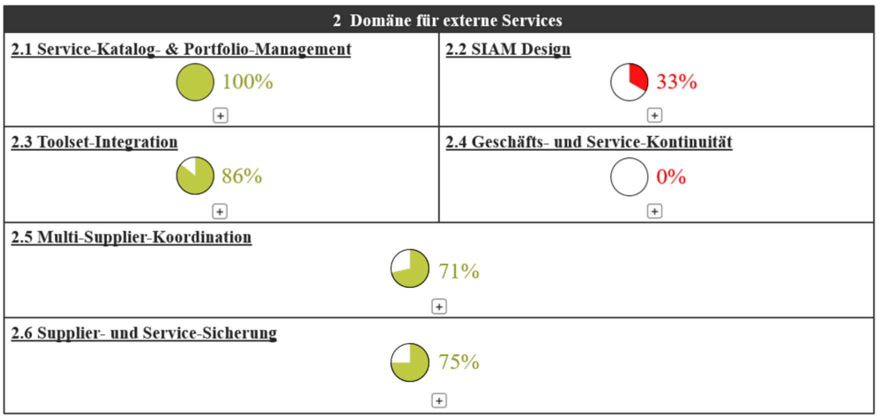


Abbildung 4. Prozessabdeckungen der SIAM-Prozesse der zweiten Domäne

In der vorliegenden Arbeit wird die Prozessabdeckung in Anlehnung an eine ITSM-Reifegrad-Studie [23] wie folgt klassifiziert: Ein SIAM-Prozess wird bei einer Prozessabdeckung unter 55 Prozent als rudimentär, zwischen 55 und 62 Prozent als mittelmäßig und über 62 Prozent als umfänglich angesehen. Die Berechnung der Prozessabdeckung dient dazu, eine übergreifende Aussage über die Abdeckung auf Prozessebene zu treffen. Die Prozessabdeckung ist allerdings mit Vorsicht zu betrachten, da sie die einzelnen Komponentenabdeckungen verschleiert und keinen Rückschluss auf die Abdeckung der einzelnen SIAM-Komponenten zulässt. Im Rahmen der Evaluation wurden die Prozessabdeckungen aller SIAM-Prozesse der zweiten Domäne ermittelt, die in Abbildung 4 dargestellt sind.

Die Übersicht über die Prozessabdeckungen lässt die Schlussfolgerung zu, dass insbesondere die SIAM-Prozesse Service-Katalog- und Portfolio-Management (2.1), Toolset-Integration (2.3), Multi-Supplier-Koordination (2.5) sowie Supplier- und Service-Sicherung (2.6) von Kingston umfänglich abgedeckt werden. Kingston ist jedoch gemäß Evaluation nur rudimentär für das SIAM-Design (2.2) und gar nicht für das Management der Geschäfts- und Service-Kontinuität (2.4) einsetzbar. Der durchschnittliche Abdeckungsgrad auf Prozessebene für die zweite Domäne liegt bei 61 Prozent. Damit ist die SIAM-Abdeckung von Kingston insgesamt als mittelmäßig einzuschätzen. Aufgrund einer Cloud-orientierten Vision von ServiceNow verfügt Kingston jedoch über ein ausbaufähiges Potenzial, um zukünftig ein umfänglich geeignetes ITSM-Tool für die Integration und das Management von IT-Services aus der Cloud zu werden.

5 Diskussion

Bei der Konzeption des Evaluationsverfahrens wurde besonderer Wert auf einfache Handhabbarkeit und geringen Anwendungsaufwand gelegt, um einen hohen Praxisnutzen zu erzielen. Um dieses Ziel zu erreichen, wurde eine gewisse Unschärfe der Bewertung in Kauf genommen. Diese resultiert zunächst aus der Verwendung von Harvey Balls in lediglich drei Ausprägungen (voller, halber, kein Harvey Ball). Die Verwendung der halb ausgefüllten Harvey Balls bedeutet nicht, dass die SIAM-Komponente durch die beiden Kingston-Funktionalitäten zu jeweils 50 Prozent und demnach zusammengefasst komplett abgedeckt wird. Ein Harvey Ball stellt keinen genauen Prozentwert dar, sondern dient der qualitativen Darstellung der Abdeckung einer SIAM-Komponente durch eine Kingston-Funktionalität. Hinzu kommt die Problematik des Skalenbruchs beim Übergang von ordinal skalierten Harvey Balls zur Repräsentation der Komponentenabdeckung hin zu kardinal skalierten Prozentwerten für die aggregierte Prozessabdeckung (vgl. [24]). Die Aussagekraft der Komponentenbewertungen ließe sich erhöhen, indem die betrachteten Tool-Funktionalitäten in einer Testumgebung unter Verwendung realitätsnaher Testdaten im Hinblick auf ihre SIAM-Konformität geprüft würden. Dadurch wäre die Verwendung von Prozentwerten auch für die Komponentenabdeckung möglich, ohne das Evaluationsverfahren substanziell zu verändern.

6 Fazit und Ausblick

Mit dem vorgestellten Evaluationsverfahren lassen sich ITSM-Tools hinsichtlich ihrer Eignung für ein integriertes Management von heterogenen Cloud-Diensten gemäß SIAM systematisch bewerten. Für das exemplarisch betrachtete Kingston wurde ein durchschnittlicher Abdeckungsgrad von 61 Prozent ermittelt, was einer mittelmäßigen Abdeckung von SIAM entspricht. Die weniger vollständigen Visionen anderer ITSM-Tool-Anbieter lassen vermuten, dass die SIAM-Abdeckung bei ITSM-Tools im Allgemeinen noch gering ist. Die starke Fähigkeit einiger Anbieter, ihre Vision umzusetzen, deutet jedoch ein großes Potenzial für eine schnelle Entwicklung der Abdeckung von ITSM-Tools an.

Um diese Folgerungen zu überprüfen, ist es erforderlich, weitere Evaluationen für ITSM-Tools durchzuführen. Auf dieser Basis wäre außerdem eine bessere Klassifizierung des Abdeckungsgrads möglich. Durch den Vergleich der Abdeckungsgrade verschiedener ITSM-Tools ließe sich besser einschätzen, ab bzw. unter welchem kumulierten Abdeckungsgrad ein ITSM-Tool reif bzw. unreif ist.

Für die durchgeführte Evaluation wurde der Fokus auf Domäne 2 des Kriterienkatalogs eingegrenzt. Um die Aussagekraft zu erhöhen, sollten bei weiteren Evaluationen ebenfalls die SIAM-Komponenten der Domänen 1 und 3 einbezogen werden. Des Weiteren erfolgte der Vergleich der SIAM-Komponenten mit den Kingston-Funktionalitäten basierend auf Beschreibungen und demnach subjektiv. Um die Analyseergebnisse zu verbessern, sollten für weitere Evaluationen objektive Bewertungskriterien und -regeln erarbeitet werden. Die praktische Erprobung vorausgewählter ITSM-Tools zur Bewertung der Abdeckung der SIAM-Komponenten ist eine Möglichkeit, eine objektivere und validere Bewertung zu realisieren.

Literatur

1. Urbach, N., Ahlemann F.: Die IT-Organisation im Wandel: Implikationen der Digitalisierung für das IT-Management. HMD Praxis der Wirtschaftsinformatik 54, S. 300-312 (2017)
2. Bitkom: Nutzung von Cloud Computing in Unternehmen boomt, https://www.bitkom.org/Presse/Presseinformation/Nutzung-von-Cloud-Computing-in-Unternehmen-boomt.html (Abgerufen: 18.06.2018)
3. Pröhl, T., Repschläger, J., Erek, K., Zarnekow, R.: IT-Servicemanagement im Cloud Computing. HMD Praxis der Wirtschaftsinformatik 49, S. 6-14 (2012)
4. Andenmatten, M.: SIAM - Das Service Integration Modell im Multiprovider Umfeld, https://blog.itil.org/2015/06/allgemein/siam-das-service-integration-model-im-multiprovider-umfeld/ (Abgerufen: 10.06.2018)
5. Matchett, C., Dohney, R., Gonzalez, K.: Magic Quadrant for IT Service Management Tools, https://www.gartner.com/doc/reprints?id=1-49P9218&ct=170811&st=sb (Abgerufen: 24.06.2018)
6. Agutter, C. et al.: Service Integration and Management Foundation Body of Knowledge (SIAM® Foundation BoK). Van Haren Publishing, Zaltbommel (2017)
7. ServiceNow: Kingston, https://docs.servicenow.com/category/kingston (Abgerufen: 19. 06.2018)

8. Holland, K: An example ITIL-based model for effective Service Integration and Management. Whitepaper. Axelos Ltd. (2015)
9. Heininger, R., Wittges, H., Krcmar, H.: Literaturrecherche zu IT-Servicemanagement im Cloud Computing. HMD Praxis der Wirtschaftsinformatik 49(6), S. 15-23 (2012)
10. Robrecht, A., Möller, C., Dörbecker, R., Böhmann, T.: Entscheidungsunterstützung für Cloud Computing – ITIL Service Lifecycle. HMD Praxis der Wirtschaftsinformatik, 49(6), S. 24-32 (2012)
11. Liebhart, D., Wischki, C.: Im IT-Sourcing ist zunehmend Flexibilität gefragt: Enorme Erleichterung. Manage IT (9-10), S. 40-41 (2009)
12. Arabalidousti, F., Nasiri, R., Davoudi, M. R.: Developing a New Architecture to Improve ITSM on Cloud computing Environment. International Journal on Cloud Computing: Services and Architecture (IJCCSA) 4(1) (2014)
13. Holland, K.: An introduction to Service Integration and Management and ITIL. Whitepaper. Axelos Ltd. (2015)
14. Nägele, D., Zander, R., Gräf, J., Auth, G.: Kritische Erfolgsfaktoren des IT-Multisourcings mit Service Integration and Management (SIAM). Tagungsband der Multikonferenz Wirtschaftsinformatik, S. 1777-1788 (2018)
15. Hallikainen, E.: Service Catalog and CMDB as the Foundation of SIAM. Master-Thesis, Haaga-Helia University of Applied Sciences, Finnland (2015), http://www.theseus.fi/handle/10024/98385 (Abgerufen: 20.06.2018)
16. Armes, D., Engelhart, N., McKenzie, P., & Wiggers, P.: SIAM: Principles and Practices for Service Integration and Management. Van Haren Publishing (2015)
17. Virri, O.: Service Integration and Management (SIAM) Practices and Four Case Studies of Finnish Companies (2017). Bachelor-Thesis, Helsinki Metropolia University of Applied Sciences, Finnland, http://www.theseus.fi/handle/10024/127271 (Abgerufen: 20.06.2018)
18. Lehtonen, V.: Practical guidelines for selecting a suitable Service Integration and Management (SIAM) model: findings from expert interviews and a case study (2017). Master-Thesis, Aalto University, Finnland, https://aaltodoc.aalto.fi/handle/123456789/25071 (Abgerufen: 20.06.2018)
19. Berghäll, J.: Service Desk meets SIAM: Integrating people, process and tools, https://www.scopism.com/service-desk-meets-siam/ (Abgerufen: 10. Juni 2018)
20. Frank, U.: Evaluation von Artefakten in der Wirtschaftsinformatik. In: Häntschel, I., Heinrich, L.J. (Hg.): Evaluation und Evaluationsforschung in der Wirtschaftsinformatik, S. 35-48. Oldenbourg, München, Wien (2000)
21. Heinrich, L.J.: Bedeutung von Evaluation und Evaluationsforschung in der Wirtschaftsinformatik. In: Häntschel, I., Heinrich, L.J. (Hg.): Evaluation und Evaluationsforschung in der Wirtschaftsinformatik, S. 7-22. Oldenbourg, München, Wien (2000)
22. Krcmar, H.: Informationsmanagement. 6. Aufl., Springer Gabler, Berlin, Heidelberg (2015)
23. ITSM Group: ITSM-Performance 2014, Bodenheim (2014), https://www.itsm-consulting.de/news-events/mediathek-downloads/leitfaeden/itsm-group_itsm-performance-studie-2014.pdf (Abgerufen: 03.12.2018)
24. Zelewski, S., Bruns, A.S.: Leitlinien für eine Case-Study-Bearbeitung aus betriebswirtschaftlicher Perspektive. In: Akca, N., Bruns, A.S., Fromen, B., Zelewski, S. (Hg.): Case-Study-Guide, Grundlagen – Anschauungsbeispiele – Hinweise für Seminararbeiten. Nomos, Berlin (2012).

How Software Promotes the Integration of Sustainability in Business Process Management

Maren Stadtländer[1], Thorsten Schoormann[1], and Ralf Knackstedt[1]

[1] University of Hildesheim, Department of Information Systems and Enterprise Modelling, Hildesheim, Germany
{maren.stadtlnder,thorsten.schoormann,ralf.knackstedt}@uni-hildesheim.de

Abstract. Business research and practice increasingly focus on integrating sustainability in organizations. To contribute to rising challenges related to the society and the environment, sustainability-driven concepts (e.g., environmental-friendly) have to be implemented in the daily business routines, and thus, need to be considered during the design of business processes in any organization. In this study, we conceptualize the field of Green Business Process Management (BPM) and use the derived concepts to classify supporting modelling tools and concrete software features. While our study indicates a lack of realization of the ecological and social sustainability in particular and a gap in supporting users during the redesign phase, there are software features that can potentially serve as a starting point to further incorporate sustainability in the design, implementation, and controlling of business processes.

Keywords: Green BPM, Classification, Tool Support, Software Feature.

1 Introduction

Due to complex challenges related to the environment (e.g., climate change) and our society (e.g., human equity) [1], sustainability has become of increasing interest in business research and practice [2, 3]. Facing these challenges, businesses need to transform themselves to allow changes in behaviour and practice as well as to fulfil new demands like being environmentally friendly [2, 4]. Therefore, novel and more sustainable business models need to be developed [5] that not only rethink separated aspects, but rather focus on radical innovations across entire businesses [6]. For implementing such new business models, it is important to align the daily, underlying business routines [7]. Therefore, on an operative level, approaches from business process management (BPM) come into play, contributing to the understanding, modelling, implementing, and optimizing business processes by, for instance, providing methods such as modelling techniques and (software) tools [8–10].

BPM is a widely used approach to develop or change organizational structures as well as increase awareness of business processes [11]. Since business processes are an essential aspect of organizations, it is necessary to properly consider sustainability-driven concepts during the design of such processes [12, 13]—it is assumed that this

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

will eventually lead to a more sustainable organization [14]. To design sustainable processes, an increasing amount of approaches for reducing energy or emissions in particular are proposed, which are typically discussed in the context of *Green BPM* [2, 15]. Green BPM takes a holistic view on the field and aims to respect sustainability across the entire lifecycle [16], for example, by integrating Green Business Process Patterns, Green Process Benchmarking, Key Ecological Indicators, or Energy-Aware Adaption (e.g., [17]). Corresponding modelling notations deal among other aspects with representing fuel or paper consuming activities [18], visualizing CO_2 footprints [19], or calculating energy flows [20] as well as support the representation of environmental impacts by integrating values and key factors for analysing the current state and optimization potential for reaching green goals [13, 15, 16]. To do so, these approaches often deal with adapting existing notations such as Business Process Model and Notation (BPMN) or Event Driven Process Chain (EPC) [16, 18, 21, 22]. Considering sustainability, according to existing literature in the field, also calls for changes to further parts of the six core elements of BPM, e.g., by using Sustainability Balanced Score Cards, adapting the process performance management to include sustainability factors, or drawing from Energy Informatics [17].

For contributing to sustainability in BPM and for applying such notations, following Recker's [23] statement that 'modelling with tools is easier', appropriate software tools need to be provided that fulfil new requirements [22]. However, although there is quite a large set of BPM software as well as different adaptions and extensions towards Green BPM, we currently lack matching these streams. To the best of our knowledge, only very few research studies (e.g., [24]) investigate some selected modelling software and compare it with criteria derived for BPM. Accordingly, this study is guided by the following key question: *How does current BPM software support the design, implementation, and monitoring of business processes in terms of economic, but primarily of ecological and social sustainability?*

For answering this question, as a first step, we conceptualized the field of Green BPM in order to derive types of adaptions, extensions, approaches, and criteria. Based on this conceptualization, we carried out an extensive software tool search and classified a selected subset of these tools. Afterwards, we analysed which software features are currently available and compared them to the derived concepts of Green BPM. Our contribution is a classification that helps academics and practitioners, for instance, to get an overview of available modelling software and to get insights on how to (re-)design current features (tool designer). Overall, we hope that our findings may also act as a starting point for more research on requirements for tools that support sustainability in businesses processes as they are important for any organization.

2 Software Tools for (Green) BPM

In this section, we outline the research background of our article in order to provide an overview of studies that already deal with conceptualizing or addressing both fields software tools for BPM and Green BPM in general.

Starting from a literature review summarizing the current state of Green BPM research [17], we identified an evaluation of supporting process modelling tools [24] and a comparative overview of process simulation tools for Green BPM [25].

Opitz et al. [24] chose BPMN as a suitable modelling language for monitoring energy efficiency KPIs as an instance of ecological process indicators and tested five of over 70 BPMN modelling tools against project-specific criteria. The highest-scoring software, ARIS Business Architect, was later used in three case studies and proved well-suited for Green BPM in the specific case of energy efficiency monitoring. The study, however, focused on the ecological dimension of sustainability, without considering social KPIs.

Lübbecke et al. [25] examined different tools for business process simulation and, upon identifying their limits, developed their own EPC-based approach for decision-making in Green BPM to support process simulation and optimization using an existing simulation tool (*Plant Simulation*) as well as energy consumption data from an external tool. To this purpose, they extended the functionality offered by the simulation tool using the built-in scripting language. However, they recognize the need for further research in order to make the data collection and modelling process easier and enable automated import of simulation data.

Riemer et al. [26] have conducted a review of BPM tools, analysing which functionality is used in order to support the (collaborative) modelling process. The tools for this review were discovered through an internet and literature search and underwent a filtering and classification process in which, starting from an initial set of criteria, each tool was evaluated. The criteria set evolved throughout this process by supplementing additional criteria and reclassifying already evaluated tools. The final set consisted of three subsets (process modelling, collaboration, and technical criteria), while the final list of evaluated software contained 11 process modelling tools. Drawing from their results, Riemer et al. [26] not only propose an extensive list of features in process modelling software focusing on collaborative modelling, but were also able to develop a high-level architecture of collaborative process modelling tool support.

3 Research Design

3.1 Conceptualizing Green BPM

As a first step, we applied a deductive approach to derive dimensions and characteristics from prior studies that are related to Green BPM. We initially focused on existing reviews and meta-analysis of Green BPM that already aim to summarize knowledge. To do so, we selected search items [27] and searched for the keywords *"green business process management"*, *"green BPM"*, and *"literature review"* in AISeL and Google Scholar (search was conducted in July 2018). By employing this search phrase, we identified 57 articles across both sources. These articles were evaluated by reading the title, abstract, and keywords. Articles that are not in the scope of this study were eliminated. Afterwards, we obtained seven articles that met our purpose of structuring different concepts of Green BPM [16, 17, 24, 28–31].

Next, we independently analysed the results to obtain fundamental concepts of Green BPM and consolidated the results to achieve a common understanding (Table 1). In the following stages, these concepts are used as a coding schema to identify features that should be implemented by software tools in the context of Green BPM.

Table 1. Consolidated concepts for Green BPM based on prior literature

Concept	*Description*	*Reference*
Notation adaptation	Adapting modelling notation elements (e.g., BPMN, UML, or EPC) for representing sustainability.	[17, 24, 28, 30, 31]
Notation extension	Extending modelling notations (e.g., adding new elements) for representing sustainability.	[17, 28, 30, 31]
Process calculation	Describing indicators and their (mathematical) relations in a formal way (e.g., to enable simulation).	[31, 32]
Process simulation	Simulating process parts and entire business process models for analysing sustainability.	[16, 17, 24]
Process optimization	Optimizing processes regarding their sustainable impact.	[30, 32]
Process benchmark	Benchmarking different processes in respect of sustainability.	[17, 29]
Process pattern	Adding patterns to apply existing knowledge to enhance processes towards sustainability.	[17, 31]
Process performance	Developed and/or adopted performance measurement methods to capture sustainable process performance.	[30]
Process indicator	Developing and analysing Green BPM indicators e.g.: *Ecological*: {energy, water, waste, resources, power usage, fuel, paper, oil, toxic} consumption; {CO_2, GHG, greenhouse, noise} emissions; recycling, air quality, renewable resources … *Social*: accidents, radiation, workforce size, health, safety, education, earnings, equity …	[17, 24, 28–32]
BPM Lifecycle Extension	Extending existing BPM lifecycles and frameworks to respect sustainability goals.	[28, 30]

3.2 Gathering BPM Software Tools

Due to the large number of available software tools for BPM, we decided to focus on tools using one specific process modelling language, namely, BPMN. Although further graphical languages such as EPC, UML, and Petri Nets exists, BPMN 2.0 is well-accepted in the context of BPM [33], and thus, selected in this study.

A Google search was conducted to reach a broad overview using combinations of the search items "*BPMN 2.0*", "*software*" (as well as the synonyms "*tool*" and "*program*"), and "*model*" (as well as derived terms such as "*modeling*/*modelling*").

Initially, after examining the first nine pages—that is, 90 individual web pages—an unevaluated list of 90 modelling tools was derived.[1]

After collecting the first results, each tool was examined. If a tool was no longer available but substituted with a new or updated version, the tool was replaced by the current software version. Tools that are part of a tool family, of which several tools allow for BPMN 2.0 modelling, were split accordingly. Software that no longer includes BPMN 2.0 modelling or was mistakenly marked as such was removed from the list. The list was cross-checked against a more extensive overview of modelling tools for various notations, leading to some additions.

The information on all tools was gathered from the software suppliers' websites and additional sources, such as product-related blogs (e.g., [34]) and studies [35, 36]. Due to financial and temporal constraints, the list for the analysis was narrowed down to 24 tools of which the majority is freely available. This choice was made because (1) academic modellers often use free software due to financial restrictions and higher perceived cost-effectiveness as well as (2) we would argue that especially (green) start-ups often lack the funds to acquire expensive software. For selection, the tools were grouped by the type of software (client, web-based, cloud-based, other, n. s.; though n. s. and cloud-based were excluded) and randomly selected in relation to the percentage of tools in each group. During the detailed analysis, *FujitsuRunMyProcesses* was removed as the current version does not support BPMN 2.0 anymore. Thus, only four shareware products remained.

3.3 Deriving Features for the Green BPM Concepts

For deriving software features, each tool was analysed regarding the concepts (cf. Section 3.1) in an inductive manner (i.e., empirically seeking for related software features). To do so, the modelling of a scenario—a simple artificial purchase and return process—guided this search, which was executed by the paper's main author. For modelling the scenario, the modeller had to carry out four main tasks, which attempt to cover all of the identified concepts: First, several sustainability performance indicators need to be included. Second, existing notation elements should be adapted and new elements need to be added. Third, the process should be monitored, simulated, automated, and analysed. Fourth, patterns and process optimization techniques need to be used. Thus, all defined concepts were covered by the case. Using the concept-based scenario, the supporting features were identified: if the software enabled fulfilling the concept—i.e., if it included a feature that could be used to model the new elements, performance indicators, or to monitor, simulate, or analyse the process etc.—the feature used was noted down. The features were accordingly clustered into groups (cf. Table 3, Section 4). On average, it took 20 minutes to examine each tool.

[1] Several institutions provide extensive lists of BPM software tools (e.g., Fraunhofer Institute or BPM&O GmbH). These lists were also included in our research.

4 Results

In Table 2, the results for each of the 23 software tools in our sample is summarized. For the classification, we particularly built on the conceptualization (cf. Section 3.1) and differentiated between the following five main dimensions:

- *Sustainability*—the first dimension differentiates between the common pillars of sustainability, namely, economic, ecological, and social (e.g., Triple Bottom Line), to indicate which type of sustainability is generally addressed by a software tool.
- *BPM lifecycle*—the second dimension makes use of the BPM phases from Dumas et al. [37] in order to visualize in which activity a software tool provides support.
- *Notation*—in the third dimension, we distinguish between adaptable (are notation elements modifiable?), extendable (can new notation elements be added?), and process patterns (included by the software manufacturer, community or modeller) to indicate how sustainability can be integrated in a modelling technique. The ability to modify or add notation elements may, among other things, indicate how much flexibility a user has to integrate their own ideas of how to depict sustainability.
- *Analysis*—the fourth dimension, for analysing business processes, differentiates between calculation (e.g., calculation of sustainability-specific variables), simulation (complete process simulation), animation (e.g., clicking through process paths), and comparison (of two or more process models, for example, through comparing simulation results). Especially simulation and comparison features may help a user to consider the ecological or social impact of their processes before implementation and enable them to compare different versions.
- *Performance indicators*—in the fifth dimension, we first aim to represent how indicators can be used by users and distinguish between usable as-is (e.g., if the tool offers an indicator), adoptable (an existing indicator may be repurposed), or extendable (new indicators may be added). This way, we may learn how difficult and/or time-consuming the use of sustainability indicators is in current tools. Further, we explored if software tools allow measuring specific ecological and social indicators found in the literature.

Moreover, to illustrate (1) how the sustainability-oriented concepts are realized by concrete software features and (2) how these are distributed in the tools (e.g., to indicate which features are often implemented as well as which features are rather neglected by software tools), we developed a morphological box (Table 3).

By using the morphological box, we aim to provide more detailed guidance on how features can be applied to contribute to sustainability in processes. Changing the colour of elements (here understood as process steps) and annotating processes (here understood as complete processes) represent easily implemented ways of highlighting (less) sustainable parts of the process. A user may, for example, decide to set the colour of process steps perceived as unsustainable to red, or add annotations (often in the form of virtual post-its or notes) describing social or ecological implications. More complex features such as calculating green KPIs during simulation or process execution, support sustainability monitoring and selecting a process with a better impact.

Table 2. Classification of software tools
(● full realization, ○ partial realization; [*] no graphical GUI, [+] simulation via third-party tools)

		Software tool/ Conceptualization	ADONIS Community Edition	AuraPortal Helium Modeler	Bonita BPM Studio Community	BPMN Visio Modeler	Camunda BPM Community Platform	Camunda Modeler	Eclipse BPMN2 Modeler[+]	Flowable[*]	GeneXus Business Process Modeler	HEFLO	Imixs-BPMN	Innovator for Business Analysts	jBPM	LucidChart	Modelio	ProcessHub	Prologics FireStart	QUAM	RunaWFE	Semtalk	ViFlow	W4 BPMN+ Web Modeler	yED Live
€		Freeware (F)	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	-	-	●	-	-	●	●
		Shareware (S)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	●	-	●	●	-	-
Sustainability		Ecological	○	-	○	○	-	-	-	○	-	-	-	○	○	-	○	-	●	n/a	○	-	○	-	-
		Economic	●	●	○	○	-	-	-	○	-	-	-	●	●	-	○	-	●	n/a	○	-	○	-	-
		Social	○	-	○	○	-	-	-	○	-	-	-	○	○	-	○	-	○	n/a	○	-	○	-	-
BPM Lifecycle		Identification	●	-	-	●	-	-	-	-	-	-	-	-	-	-	○	-	●	●	-	-	●	-	-
		Discovery	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
		Analysis	●	●	●	-	-	-	○	-	-	●	-	○	●	-	-	-	●	n/a	●	●	n/a	-	-
		Redesign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	-	-	n/a	-	-	-	-	-
		Implementation	-	-	●	-	●	-	-	●	●	●	●	-	●	-	-	-	●	n/a	-	-	n/a	-	-
		Monitoring	-	-	●	-	●	-	-	-	n/a	●	●	-	●	-	-	-	●	●	-	-	n/a	-	-
		Extendable	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	n/a	-	-	-	-	-
Notation		Adapt	-	-	-	-	-	-	-	-	-	-	●	-	-	-	-	-	-	n/a	-	-	-	-	-
		Extend	●	-	●	●	-	-	-	-	-	-	-	○	○	●	●	-	●	●	○	-	○	-	●
		Process pattern	-	-	-	-	-	-	●	-	-	-	●	-	-	-	●	-	-	n/a	-	-	-	-	-
Analysis		Calculation	●	●	●	-	-	-	-	-	-	-	-	●	●	-	-	-	○	n/a	●	-	○	-	-
		Simulation	●	●	-	-	-	-	○	-	-	●	-	○	●	-	-	-	-	n/a	●	●	n/a	-	-
		Animation	-	-	-	-	-	-	-	-	-	-	-	●	-	-	-	-	-	n/a	-	-	-	●	-
		Comparison	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	n/a	-	-	-	-	-
Performance Indicator	Use	Useable	●	●	-	-	-	-	-	-	-	-	-	-	●	-	-	-	●	n/a	-	-	○	-	-
		Adoptable	●	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	n/a	-	-	-	-	-
		Extendable	-	-	●	●	-	-	-	-	-	-	-	●	●	-	-	-	●	○	○	-	●	-	-
	Ecological	Paper con.	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		Energy con.	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		CO_2 emission	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		Other	○	-	○	○	-	-	-	○	-	-	-	○	○	-	-	-	●	n/a	○	-	○	-	-
	Social	Employee health	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		Employee equity	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		Employee security	○	-	○	○	-	-	-	○	-	-	-	-	○	-	-	-	○	n/a	○	-	○	-	-
		Other	○	-	○	○	-	-	-	○	-	-	-	○	○	-	-	-	○	n/a	○	-	○	-	-

Table 3. Features for implementing the Green BPM-concepts

Concept	*Software featuress*				
Notation adaptation	changing colour of elements (10/23)	adapting BPMN tasks/events to trigger automated reports (1/23)	linking elements with other processes/other resources[1] (6/23)	annotating[1] (12/23)	
Notation extension	adding new elements (4/23)	adding new business objects/process variables (4/23)	adapting elements from elements of other notations (7/23)	adding views (1/23)	
Process calculation	calculating via simulation (5/23)	calculating during process execution (3/23)	generating reports (2/23)	calculating single variables (3/23)	calculating all variables (5/23)
Process simulation	using own simulation engine (7/23)	using external simulation plugin (1/23)	simulating to generate KPIs/ variables (6/23)	checking correctness via simulation[1] (1+/23)[2]	simulating reports (1/23)
Process optimization	using built-in comparison (1/23)	using built-in improvement feature (1/23)	importing customized process pattern catalogues (1/23)		
Process benchmark	using built-in comparison (1/23)	using external comparison (0/23)			
Process indicator	using implemented KPIs (5/23)	extending a set of business/process variables/KPIs: calculable (6/23)	extending a set of business/process variables/KPIs: non-calculable (2/23)	adopting risk models (3/23)	
Legend:	0	1-4	5-8	9-12	>12

5 Discussion and Implications

5.1 Describing the Current State of Software Tools for BPMN

Based on our analysis, the following main observations emerged: First, while most tools in our sample focus on *discovery* and *analysis*, the phase of *redesign* in the BPM lifecycle is rather neglected in current BPMN software, which complicates revising existing business processes towards sustainability. For instance, users may compare two versions of a process diagram—manually or by comparing simulation results—but only one software offers an option for redesign by allowing to include a library of process patterns. The process patterns included in other tools only consist of general workflow patterns, intended rather for saving time while modelling than for redesign.

Second, for visualization and increasing the awareness of sustainability-oriented aspects in current business processes, a number of tools allow applying elements of other notations or self-created elements. Extending the notation can be carried out by modelling the new elements in the BMPN 2.0 diagram itself, or by referencing between processes or process elements and elements within diagrams of other types. However, repurposing or creating elements defined by single users and not the software provider

1 Adapted from Riemer et al. [26].

2 As this feature was not part of our analysis criteria and the modelled example process, we cannot give the exact number of tools implementing it.

themselves, would most likely not be shared with a broader community of business process modellers, thus encouraging isolated (and heterogeneous) application.

Third, sustainability indicators are rarely implemented in our sample, with economic indicators (e.g., costs and time) outweighing ecological and social ones. This is surprising because in contrast to the tools research exists that already explores, for example, key ecological indicators in the context of BPM (e.g., [16]). The tools that contain indicators for costs use pre-set variables, often dubbed as 'resource KPI', 'KPI', or 'business variable'. Only *Prologics FireStart* includes one pre-set ecological indicator ('Umweltindex') in the form of a (unspecified) scale from 1 to 10, which leaves the interpretation open for users. Some other tools allow adding own indicators by using variables referencing risk model elements (including ecological, economic, and social risks, as defined by the modeller). In some cases, where no indicators for social or ecological sustainability exist by default, features are incorporated that may be repurposed to act as indicators: most notably risk models, user-created process variables, objects, or KPIs, or analysable extension fields. These would need to be explicitly created and adapted by the modeller. While this adaptation and repurposing enables, for example, the highlighting of fuel and paper consuming activities [18] or calculation of energy flows [20] as described in the literature, the lack of pre-installed extensions and examples for sustainability indicators may hinder reflection of ecological or social aspects within the modelling process.

5.2 Deriving a Preliminary Set of Requirements for BPM Software Tools

Based on our findings from exploring a sample of current software tools for BPM as well as comparing them with sustainability-oriented concepts that are grounded in previous literature, we aim to specify a first set of (new) requirements for such tools:

- *Integrated catalogue of Green BP Patterns*—Implementing a pre-installed catalogue of green patterns in order to support users to, for example, identify potential for improvement of current processes as well as get ideas on how to improve current deficits in terms of ecological and social sustainability (e.g., [25, 27, 38]).
- *Sustainability-oriented performance indicators*—Including pre-installed indicators for ecological and social sustainability in order to facilitate, for example, the measurement of business process outcomes (e.g., as researched in [17, 24, 28–32]).
- *Comparison of process variants*—Providing feature(s) for comparing the ecological, economic, and social impact of different processes or process variants in order to support the users in making informed decisions for the redesign.
- *Data import*—Implementing interfaces for (automatic) import of data related to sustainability-oriented indicators, for instance from sources like energy monitoring systems, to support both process simulation and real-time monitoring.
- *Sustainability-driven constructs*—Providing notations to visualize ecological (e.g., resource consumption) and social aspects (e.g., workplace conditions) in processes.

Of course, this set of requirements for software tools supporting Green GPM needs to be evaluated in future steps, for example, by building prototypes that instantiate derived requirements and demonstrate the utility of them (e.g., [39]).

6 Conclusion

In this study, we first conceptualized Green BPM and classified a sample of 23 existing BPMN 2.0 software tools in order to analyse the degree of satisfying the needs of Green BPM. Our results have yielded gaps in the support of Green BPM, especially in terms of social and ecological sustainability. Based on our results, we formulated a first set of requirements for such a class of software tools that need to be evaluated in future research. The analysis may help users who want to focus on the sustainability of their processes over the course of the BPM lifecycle to get an overview of suitable modelling software or how to reuse certain existing features for this purpose. Furthermore, it offers a guidance for software designers and researchers on where to begin enhancing existing software for Green BPM or designing altogether more sustainable modelling software.

Although we derived helpful insights for the field of Green BPM, this study is not free of limitations. First, the analysis mostly covers freeware tools, which may have a limited range of features compared to shareware software. Nevertheless, we aimed to select a representative set of tools as well as would argue that, for instance, start-ups with limited financial resources addressing sustainability might need freeware tools to support the operationalization of their ideas. Second, although we specified relevant concepts that should be implemented in software by reviewing previous literature, there might be more aspects that need to be considered in such tools. Third, the identification of corresponding features is based on own interpretations and decisions, which also have limitations (e.g., other researchers might identify more software features).

In future research, we plan to expand and evaluate the list of modelling software as well as the preliminary set of requirements. Using our analysis as a first step, it is now possible to generate guidelines for designing and developing BPM software that helps users examine ecological and social aspects and to examine the potential and necessity of implementing further sustainability indicators (e.g., as suggested by the Global Reporting Initiative). This is a necessary step in order to introduce a proactive consideration of multidimensional sustainability into the routine of organizations.

References

1. Brundtland, G.H.: World Commission on Environment and Development: Our common future. Oxford: Oxford University Press (1987)
2. Brocke, J. vom, Seidel, S., Recker, J.: Green Business Process Management: Towards the Sustainable Enterprise. Springer Science & Business Media (2012)
3. Melville, N.P.: Information Systems Innovation for Environmental Sustainability. MISQ 34, 1–21 (2010)
4. Hanelt, A., Busse, S., Kolbe, L.M.: Driving business transformation toward sustainability: exploring the impact of supporting IS on the performance contribution of eco-innovations. Inf. Syst. J. 27, 463–502 (2017)
5. Schaltegger, S., Freund, F.L., Hansen, E.G.: Business cases for sustainability: the role of business model innovation for corporate sustainability. Int. J. Innov. Sustain. Dev. 6, 95 (2012)

6. Nidumolu, R., Prahalad, C.K., Rangaswami, M.R.: Why Sustainability Is Now the Key Driver of Innovation. Harv. Bus. Rev. (2009)
7. Al-Debi, M.M., El-Haddadeh, R., Avison, D.: Defining the business model in the new world of digital business. In: Proc. of the Americas Conference on Information Systems, Toronto, Canada (2008)
8. Hammer, M.: What is Business Process Management? In: Brocke, J. vom and Rosemann, M. (eds.) Handbook on Business Process Management 1. pp. 3–16. Springer Berlin Heidelberg (2010)
9. Rosemann, M., De Bruin, T.: Application of a holistic model for determining BPM maturity. BP Trends. 1–21 (2005)
10. Scheer, A.-W., Brabänder, E.: The Process of Business Process Management. In: vom Brocke, J. and Rosemann, M. (eds.) Handbook on Business Process Management 2. pp. 239–265. Springer Berlin Heidelberg (2010)
11. Rosemann, M.: Potential pitfalls of process modeling: part A. Bus. Process Manag. J. 12, 249–254 (2006)
12. Loos, P., Nebel, W., Marx Gómez, J., Hasan, H., Watson, R.T., vom Brocke, J., Seidel, S., Recker, J.: Green IT: A Matter of Business and Information Systems Engineering? Bus. Inf. Syst. Eng. 3, 245–252 (2011)
13. Seidel, S., vom Brocke, J., Recker, J.C.: Call for action: investigating the role of business process management in green IS. In: Hasan, H. and Dwyer, C. (eds.) Sprouts: Working Papers on Information Systems. AIS, Saint Louis, Missouri (2011)
14. Levina, O.: Exploring the Role of Business Process Management in Sustainability Initiatives. In: Proc. of the Mediterranean Conference on Information Systems, Samos, Greece (2015)
15. Meyer, J., Teuteberg, F.: Nachhaltiges Geschäftsprozessmanagement–Status Quo und Forschungsagenda. In: Proc. of the Multikonferenz der Wirtschaftsinformatik, pp. 1515–1530. Braunschweig, Germany (2012)
16. Opitz, N., Krüp, H., Kolbe, L.M.: Environmentally Sustainable Business Process Management-Developing a Green BPM Readiness Model. In: Proc. of the Pacific Asian Conference on Information Systems (2014)
17. Maciel, J.C.: The Core Capabilities of Green Business Process Management–A Literature Review. In: Proc. of the Wirtschaftsinformatik, pp. 1526–1537. St. Gallen, CH (2017)
18. Recker, J., Rosemann, M., Hjalmarsson, A., Lind, M.: Modeling and Analyzing the Carbon Footprint of Business Processes. In: vom Brocke, J., Seidel, S., and Recker, J. (eds.) Green Business Process Management, pp. 93–109. Springer Berlin Heidelberg (2012)
19. Houy, C., Reiter, M., Fettke, P., Loos, P., Hoesch-Klohe, K., Ghose, A.: Advancing Business Process Technology for Humanity: Opportunities and Challenges of Green BPM for Sustainable Business Activities. In: Brocke, J. vom, Seidel, S., and Recker, J. (eds.) Green Business Process Management, pp. 75–92. Springer Berlin Heidelberg (2012)
20. Hoesch-Klohe, K., Ghose, A.: Carbon-Aware Business Process Design in Abnoba. In: Maglio, P.P., Weske, M., Yang, J., and Fantinato, M. (eds.) Service-Oriented Computing, pp. 551–556. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
21. von Ahsen, A.: Qualitäts-und umweltorientierte Ausgestaltung von Führungs-und Leistungsprozessen. Integriertes Qualitäts-und Umweltmanagement: Mehrdimensionale Modellierung und Umsetzung in der deutschen Automobilindustrie (2006)
22. Houy, C., Reiter, M., Fettke, P., Loos, P.: Towards Green BPM – Sustainability and Resource Efficiency through Business Process Management. In: Business Process Management Workshops, pp. 501–510. Springer, Berlin, Heidelberg (2010)

23. Recker, J.: “Modeling with tools is easier, believe me”—The effects of tool functionality on modeling grammar usage beliefs. Inf. Sys. 37, 213–226 (2012)
24. Opitz, N., Erek, K., Langkau, T., Kolbe, L., Zarnekow, R.: Kick-starting Green Business Process Management–Suitable Modeling Languages and Key Processes for Green Perfomance Mesaurement. In: Proc. of the American Conference on Information Systems. Seattle, USA (2012)
25. Lübbecke, P., Reiter, M., Fettke, P., Loos, P.: Simulation-Based Decision Support for the Reduction of the Energy Consumption of Complex Business Processes. In: Proc. of the Hawaii International Conference on System Science, pp. 866–875. IEEE (2015)
26. Riemer, K., Holler, J., Indulska, M.: Collaborative process modelling-tool analysis and design implications. In: Proc. of the European Conference on Information Systems, p. 39. Helsinki, Finland (2011)
27. Schoormann, T., Behrens, D., Fellmann, M., Knackstedt, R.: <<Sorry, too much Information>> Design Principles for Supporting Rigorous Search Strategies in Literature Reviews. In: Proc. of the IEEE International Conference on Business Informatics. IEEE, Vienna, Austria (2018)
28. Stolze, C., Semmler, G., Thomas, O.: Sustainability in business process management research–a literature review. In: Proc. of the American Conference on Information Systems. Seattle, USA (2012)
29. Gräuler, M., Teuteberg, F.: Experimental Evaluation of a Process Benchmarking Tool in a Green Business Process Management Context. In: Proc. of the Wirtschaftsinformatik. Leipzig, Germany (2013)
30. Gohar, S.R., Indulska, M.: Business Process Management: Saving the Planet? In: Proc. of the Australasian Conference on Information Systems, pp. 1–14. Adelaide, Australia (2015)
31. Schoormann, T., Behrens, D., Knackstedt, R.: Sustainability in Business Process Models: A Taxonomy-Driven Approach to Synthesize Knowledge and Structure the Field. In: Proc. of the International Conference on Information Systems. Seoul, Korea (2017)
32. Opitz, N., Krüp, H., Kolbe, L.M.: Green Business Process Management -- A Definition and Research Framework. In: Proc. of the Hawaii International Conference on System Science, pp. 3808–3817. IEEE, Waikoloa, HI, USA (2014)
33. Harmon, P.: The state of business process management 2016. Bus. Proc. Trends (2016)
34. Hesse, M.: BPMN Tool Matrix, https://bpmnmatrix.github.io/ (Accessed: 22.09.2018)
35. BPM Toolmarktmonitor: Software für das unternehmensweite Prozessmanagement, https://bpm-expo.com/publikationen/software-fuer-das-unternehmensweite-prozessmanagement/ (Accessed: 22.09.2018)
36. Drawehn, J., Kochanowski, M., Kötter, F.: Business Process Management Tools 2014. Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO, Stuttgart, Germany
37. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer Berlin Heidelberg (2013)
38. Nowak, A., Leymann, F., Schleicher, D., Schumm, D., Wagner, S.: Green Business Process Patterns. In: Proc. of the 18th Conference on Pattern Languages of Programs. pp. 6:1–6:10. ACM, New York, NY, USA (2011)
39. Peffers, K., Rothenberger, M., Tuunanen, T., Vaezi, R.: Design Science Research Evaluation. In: Design Science Research in Information Systems. Advances in Theory and Practice, pp. 398–410. Springer Berlin Heidelberg (2012)

Criteria Catalog for Industrial IoT Platforms from the Perspective of the Machine Tool Industry

Philipp Werner[1], Dimitri Petrik[1,2]

[1] University of Stuttgart, Chair of Information Systems II, Stuttgart, Germany
{philipp.werner,dimitri.petrik}@bwi.uni-stuttgart.de
[2] University of Stuttgart, Graduate School of Excellence advanced Manufacturing Engineering (GSaME), Stuttgart, Germany

Abstract. Through providing digital services, machine tool manufacturing companies can address the customer demands for individual solutions and the increasing cost pressure. Applications are one way to provide these digital services, running on smart machine tools connected to software systems, known as industrial Internet of Things (iIoT) platforms. Despite the growing potential of iIoT platforms in the provision of industrial digital services and the increasing awareness of the platform approach among manufacturing companies, lack of requirements makes the platform challengeable for machine tool companies. Moreover, the domain specific industrial application of platforms has been limitedly researched, indicating a possible research gap. This paper presents a literature-based research on requirements for iIoT platforms, followed by the solution-oriented metrics, to fulfill each requirement. Together, the requirements and metrics form a structured criteria catalog for iIoT platforms, which can be used as a decision support tool for the machine tool industry.

Keywords: Platform Ecosystem, industrial IoT Platform, IoT Ecosystem, Criteria Catalog, Smart Machine Tool.

1 Introduction

1.1 Research Gap and industrial problem setting

Current machine tool manufacturing companies are challenged by the increasing product variety, offered by original equipment manufacturers (OEM) in different industries, and the related flexibility for their own products [1]. Customer demand for custom solutions, as well as the increasing role of the after-sales services in the competition, were also surveyed for the mechanical engineering industry by the German Mechanical Engineering Industry Association (VDMA) and McKinsey [2]. A machine tool manufacturer could address these challenges by offering digital or smart services within its service portfolio. Current empirical studies show a high interest within the machine tool industry to offer digital services to monitor machine data, in order to improve manufacturing processes [2-3]. Smart machine tools are estimated to increase machine productivity and life by about 5%, reduce maintenance costs by 10-

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

40% and reduce the energy consumption by 20% [4]. Such smart and connected machine tools are equipped with embedded computers and networks, monitoring and affecting physical processes to achieve the improvements mentioned above, through digital applications [5]. The possibility to increase the flexibility turns smart machine tools into platforms, being simultaneously a product (machine tool) and a platform (modular extension through applications) [6]. Smart machine tools are also closely related to the concept of cyber-physical systems (CPS), as they can be controlled remotely and communicate with surrounding systems and the physical world [7].

However, as various case studies show, various companies have failed to establish successful platforms in the past [8]. Current state of the market on iIoT platforms pictures its highly fragmented state, thus including more than 450 iIoT platforms available on the market [9]. In contrast to the current state of the market for iIoT platforms, platform-based markets are rather affected by the "winner-takes-all" competition logic of a platform-providing keystone company [10]. Due to the high fragmentation of the market, the right choice and utilization of an iIoT platform is a challenging task for machine tool companies and represents an entry-barrier for a provision of platform-based digital services in the after-sales. The practical problem from the perspective of the collaborative customer in the iIoT lies in the complexity during the platform selection process. On the other hand, the variety of iIoT platforms makes it difficult for the iIoT platform providers to gain significant market share, becoming a "platform leader" [11]. From the research perspective, previous scientific work on platforms focuses primarily on business-to-consumer (B2C) industries and the information technology (IT) cases [12-13], while the consideration of specific industrial requirements in the area of platforms still need further research. Additionally, the empirical studies also highlight the fact, that providing smart services through digital applications is a challenge for machine tool manufacturing companies, since 59% of those surveyed do not offer market ready solutions [2]. Considering these factors, there is an apparent need for research on these platforms for smart connected machine tools.

1.2 Research Questions and Structure

The main goal of this paper is to provide a criteria catalog for iIoT platforms as a decision support tool in the domain of smart connected machine tools, based on the definition of two research questions:

- RQ1: What are the relevant platform criteria for smart machine tools?
- RQ2: Which metrics fulfill the criteria identified in RQ1?

The second section of this paper describes major concepts and research streams, related to iIoT platforms. The first research question is addressed in the third section. Firstly, the research design of the conducted literature analysis is described, followed by the extraction of the relevant criteria. The section after that focuses on the relevant metrics of each identified criterion. These metrics provide a more detailed description of each criterion and allow a systematical application of the criteria for instance during a selection process of an iIoT platform from the perspective of a mechanical engineering company. Together, the criteria and the assigned metrics are building a structured criteria catalog with relevant criteria for an application of the platform approach for smart machine tools. The criteria catalog builds the artefact of this paper. The scientific approach to achieve the first artefact is a structured literature analysis.

The overall scientific contribution of this paper is an extension of platform research for the industrial application. The focus on smart machine tools is justified due to the high importance of the industrial IoT in this industry and could be adjusted to other manufacturing industries.

2 Theoretical background and related concepts

This contribution relies on three major concepts related to the iIoT platforms, which are explained in this section: CPS, iIoT and technological platforms.

As mentioned in the introduction, the provision of customized solutions requires machine tool companies to extract data from machine tools and use this data to control them. This is achieved due to sensors, actuators, embedded computational power and connection of a machine tool to external data-processing platforms, thus creating loops between the physical assets (machine tools) and their digital counterparts [7, 14]. The analyzed machine data can be used for autonomous parameter change of the machine by the actuators. This goes beyond the traditional automation technology of machine tools. Hence, such machine tools are defined as smart and fit the definition of CPS.

The integration of CPS in machine tools is also related to the concept of iIoT. IoT is a paradigm of the integration of internet and communication technologies (ICT) and real-time data analytics in physical assets. The application of IoT on industrial assets in the manufacturing builds the concept of iIoT [15]. Even though, IoT can be applied in various industries [16], this paper is domain specific and is based on the application of IoT in the machine tool industry. Smart machine tools, in the context of iIoT (beyond the simple automation) are able to track various sensor data, monitoring the machine itself, the manufacturing process and the quality of the manufactured product [17] and this data can be used for digital services. However, the historical evaluation of sensor-based data and the provision of the analyzed data as a service to external companies requires additional software systems [18], known as iIoT platforms.

The data analysis and related visualization are hosted as platform-based applications, which can be developed by third-party companies. Therefore, iIoT platforms are related to the technology platforms concept, defined by Cusumano and Gawer [11] and platform-based innovation by the ecosystem [13]. This concept builds the third related theoretical concept of this contribution. According to this definition, technological platforms provide a dominant technology, including industry standards and transforming them into a complete customer solution. The provider of a technological platform is considered as a keystone company [19]. Due to the control over the platform, the platform-providing keystone company is enabled to form alliances and partnerships with third-party companies, which are considered as competing complementors, consequently building a platform-based ecosystem [13]. The participation of third-party companies is supported by the modular design of technological platforms, which in the case of iIoT is represented by platform-based applications. Those modular applications extend the value of the platform through network effects [20]: if for instance an application for cost reducing condition monitoring of a machine tool is presented in the application store of an iIoT platform, more machine tool companies would consider using the platform. Similarly, more developers would complement to the iIoT platform, if a platform-based ecosystems contains enough application customers. In order to enable such a complementary third-party development of complementary modules, the interfaces of the platform have to be open. The platform provider is in the position to determine how to share the interface specifications with the third-party companies. Empirical examples of iIoT platforms, as Mindpshere (Siemens) or Thingworx (PTC) correspond to the both classifications of open technological platforms, providing specification and tools for the integration of application by complementary third-party companies [21-22]. In addition, the application stores of both iIoT platforms show exemplary applications, provided by third-party companies [21-22]. Therefore, iIoT platforms fit into the theoretical concept of open technological platforms.

3 Creating the Platform Criteria Catalog for Machine Tools

3.1 Research Methodology

The aim of the literature review is the identification of the current state of research towards the derivation of relevant criteria for platforms for machine tools based on the found literature. The structured literature research approach is sophisticated, reproducible, systematic, transparent and scientific [23-24].

The methodological approach used in this work follows the scientific work of Webster and Watson [25]. In addition, the approach is based on the work of Rashman et al., Soni and Kodali [26-27] and Winter and Knemeyer [28]. The review period for the literature review is between January 2002 and December 2017, thus covering 15 years. The reason for choosing the year 2002 as the start point is that Gawer and Cusumano distinguish different platform types in their work and subdivide platforms into internal platforms, supply chain platforms and industry platforms in that year [11]. As stated above the platforms are examined in an industrial application context. For this reason, the focus of the platform analysis is the current state of practice in regard to the classification of the platforms according [11]. Therefore, the start date can be set to 2002, since it is ensured that the term industry platform does not exist before 2002. The reason for choosing December 2017 as the end date is that the latest scientific publications should be considered in order to develop the broadest possible set of criteria and this was the current date during the review period.

Three different databases are used to conduct the structured literature research: Business Source Premier (EBSCOhost), IEEE Xplore and Science Direct. Google Scholar is also included in the research to get a broader overview of existing relevant articles and include relevant papers, which are not covered by the three databases. Business Source Premier includes a large number of business, finance and management journals and is therefore selected. To cover the technical area in relation to CPS and the industrial context, Science Direct, which includes articles on technical and engineering subjects, is also included in the research. IEEE Xplore provides articles on the world's most cited publications in electrical engineering, computer science and electronics and completes the research. The databases are searched using the following terms: "Industry Platform", "IoT Platform", "Service Ecosystem", "Service Platform" and "Software Ecosystem"; linked by the Boolean operator AND with the following terms: "Cyber Physical System", "Industrial Internet of Things", "Industry 4.0", "Machine Tools" and "Smart Manufacturing". Only English-language articles are considered.

3.2 Results from the Literature

The review process returns 147 articles after the removal of duplicates. For this purpose, the title, the abstract and the conclusion are read first from each of the 147 articles. Looking at the titles, abstracts, and conclusions, many of the articles found do not address RQ1. Therefore, 125 articles are excluded, leaving a sample of 22 articles. The backward search on the related work lists of the identified 22 articles, followed by the forward search using Google Scholar, identifies six additional articles. The sample therefore contains 28 articles, which are completely analyzed. After reading through the articles completely, another seven articles can be excluded because their contribution cannot be used for RQ1. In the end, 21 articles provide possible relevant examination criteria for intelligent machine tool platforms. However, the search includes articles which do not explicitly refer to embedded systems in the machine tool industry, but provide general criteria, which can be relevant for open platforms for intelligent machine tools. The following matrix, as proposed by Webster and Watson, depicts the found criteria and the relevant articles.

The examined articles describe criteria for intelligent machine tools highlighting various perspectives on this field. These perspectives can be classified as CPS in manufacturing, Software-Product-Service-Systems, industrial IoT Platforms, service platforms for machine tools or they offer a generic view on platforms or software ecosystems. After the detailed analysis of the described concepts in the literature, 13 identified requirements are grouped in six criteria for the machine tool platform approach, which are depicted in the following table:

Table 1. Platform criteria list for intelligent machine tools

ID	*Criterion*	*Sources*
1	Security of the platform	[29-33], [36], [39-40], [43-49]
2	Modularity of the platform	[30], [32-37], [41-42], [45], [47-49]
3	Degree of openness (Platform and Interfaces)	[29-30], [32-35], [37-38], [42-44]
4	Functionalities on the platform	[29-30], [32-33], [36], [39-40], [45], [47]
5	Range of the platform	[30], [32], [36]
6	Autonomy of all stakeholders	[35], [37]

- **Security of the platform:** Data security, information security, system security and quality assurance measures could be summed up to a key criterion “Security of the platform". It seems to be currently the most commonly cited criterion in the literature (15 of the 21 articles). Some articles describe the aspects of safety and protection of sensitive production data as the most urgent challenge for platforms [40, 48] and some consider the data processing security of CPS as an important challenge [45]. Quality assurance as a criterion, which was also cited in six articles, additionally contributes to the safety of the platform. If the open platform brings various companies (for example a machine tool company and a third-party software company) quality assurance plays an important role. If a platform could internally assist the machine tool manufacturer in achieving quality measures on third party applications, it could increase the trust of the machine tool manufacturer in the externally developed software modules. System safety means the security of the machine to its environment and ensures no persons or work pieces are harmed by incorrectly programmed software. A failure could result in serious material or personal damage [29, 44]. However, as the literature mostly states the data security and the information security, the system security will be summed up under the generic term "Security of the platform".

- **Modularity of the Platform:** Modularity also increases the scalability of a platform, which is why both terms are summed up as in the criterion a "Modularity of the Platform". Modularity seems to be an important subject of research, as it is mentioned in 13 articles. The scalability of the platform does not currently seem to attract much attention, being stated in just one article. In order to develop a uniform knowledge base for an intelligent machine tool, various companies collaborate on the open platform across numerous corporate boundaries [30, 32, 36]. The modularity of an open platform for intelligent machine tool is necessary to use the different capabilities of the involved actors as third-party software developers or sensor manufacturers and to reduce the complexity for the machine tool manufacturer. It is also discussed that platforms enable synergies between all those involved companies and can lead to more innovations [34]. Though modularity leads to a reduction of the technological complexity of the intelligent machine tool, it simultaneously increases the management and governance complexity of the platform, which can also be a major challenge for a machine tool company, as it requires additional management efforts. Finally, the intelligent machine tool has to be manageable despite its modularity and the resulting platform complexity [34]. An intelligent machine tool produces great amounts of data that has to be stored, processed and analyzed, which could present a major challenge for a manufacturer. Consequently, the platform should handle this challenge and deliver scalable resources [36]. Moreover, it becomes important for the manufacturing company to decide which modules should be developed on his own and which should be outsourced to third party companies.
- **Degree of Openness (Platform and Interfaces):** This criterion sums up openness and the integration of the platform. During the life cycle of a platform, the degree of openness can change. It could make sense to keep the platform more closed in the beginning, until it gathers a certain amount of early adopting users and developers and open it afterwards. For a mature platform with a large group of developers and users it is preferable to increase the openness, as it generates more overall value for

all involved stakeholders, through increased innovation [44]. A high degree of openness with uniform standards helps to improve the efficiency of the integration of the intelligent machine tool as a platform in the systems operated by the customer [29, 32] (for instance manufacturing execution systems or other information systems). Moreover, it helps to access all the capabilities, know-how and data with a platform-based ecosystem with a smaller investment [36-38]. In order to attract a large number of complementing third-party companies, it is recommended to provide good documentation and interfaces for the complementors [30]. A lower degree of openness requires a higher beginning invest, though it can result in a competitive advantage through the full control of development modules of a platform [38] and preventing the unsuitable complementors to access the platform [30]. Additionally, empirical analyses show the dominant usage of closed platforms by manufacturing companies, preferring to restrict the degree of openness of the platform and to collaborate only with specific partners [35, 42].

- **Functionalities on the Platform:** Functionalities unite the user interface (UI), the functionality, support and services and the test environment and test access. According to the UI as a criterion the platform services and the machine tool interface should be understandable and easy to use for the user [30]. It is also recommended to provide the machine data analysis in the browser, making it accessible with mobile devices, in order to improve the UI [39, 45]. Functionality represents the variety of analysis options and applications for the customer. The support and services target two groups. On the one hand, the customer of smart connected machine tools could require support regarding the maintenance of the intelligence of smart connected machine tools [32]. On the other hand, the complementing third-party companies could require support of a platform provider during the development. A test environment is important for the customer of a smart machine to test, whether the platform sufficiently supports the required functionality, consequently reducing the risk of bad investment and sunken costs. The complementing third-party software company could also test their prototypes and reduce their risks and training costs [29-30].
- **Range of the Platform:** The range increases the awareness level of the platform among the customers and complementing third-party companies, which indicated to what extent the platform can prevail in the future as a standard. Uniform interfaces and industry standards are required to achieve high range of a platform, thus increasing the platform complexity for the platform provider [30, 36]. These uniform standards can only be enforced if the platform has a certain reach and market share. Support in terms of quality assurance by the keystone also increases the range [30].
- **Autonomy of all Stakeholders:** All participating stakeholders are autonomous to the platform and can independently decide, whether and how they participate in the platform ecosystem. This boosts the innovative strength within the platform, due to the resulting competition [37]. If not prevented, autonomous stakeholders will either complement or even compete with their complementary innovation with the platform provider. However, high degree of autonomy of the stakeholders, increases the competition and results in higher degree of platform-based innovation [37]. Furthermore, the positively perceived autonomy of the stakeholders, can be used by

the keystone and by the complementing stakeholders to attract new stakeholders in the platform-based ecosystem, because this indicates lower dependency risks for the stakeholders [35].

3.3 Matching Requirements and Metrics to Create a Criteria Catalog

The platform provider should fulfill all the identified criteria. Hence, this section provides metrics for each criterion. This paper uses the term metrics to describe solutions, currently offered by the platform providers, in order to fulfill the requirements for a platform in the area of smart machine tools. The criteria with their metrics form the criteria catalog, thus answering the RQ2 and representing the artefact of this contribution. The identification of the metrics is based on two relevant articles [8, 50], which were congruently identified through the literature research. Both articles build the foundation for the application of solution-oriented metrics on the identified criteria for two reasons. Firstly, both articles use a similar approach, separating requirements and solutions from each other. This approach matches with a core principle in the quality function deployment (QFD). QFD separates neutral requirements from the product-specific solutions. The application of the core principle of QFD is justified by the higher efficiency of this of this method during the product development in terms of customer satisfaction [51], which seems suitable considering the high fragmentation state of the market for iIoT platforms. Furthermore, the application of QFD on software design in the past [51]. Secondly, both articles list comparable platform criteria to those identified in this paper. The following table presents the criteria catalog with the metrics, which are suited to fulfill each criterion.

Table 2. Criteria catalog including metrics

ID	*Criterion*	*Metrics*
1	Security of the platform	Backup & Recovery Data encryption Threat prevention Traceability Management of access roles and rights Single-Sign-On Existing evaluation possibilities Existing certifications
2	Modularity of the platform	Scoping the ratio between own and external development Modular machine tool design (hardware) Specification of design principles
3	Degree of openness (platform and interfaces)	APIs for expanding the connected smart machine tool Functional description of the APIs Integration possibility of non-platform machine tools Technical description of the APIs Support of various formats for data exchange Presence of standards

		Accessibility of the platform
4	Functionalities on the platform	Uniform user interface
		Social media connection
		Systems for functional testing possibilities
		Test access to the platform or a demo version
		Trainings and Certifications
		Support and transaction processing by the platform
		Availability of a help desk for complementors
		Cooperation in development
5	Range of the platform	Number of end users
		Number of complementors and partners
6	Autonomy of all stakeholders	Autonomy of complementors

4 Conclusions and Further Research Outlook

The overall benefit of this paper is a structured criteria catalog, suitable for researchers and practitioners to benchmark the highly dynamic market for iIoT platforms, identify relevant concrete criteria of a platform a improve the platform selection process for a company, connecting its assets as the smart machine tools with the platform. Moreover, the criteria could also address the platform providing companies. The business model development of a platform providing company for instance could use the results to uncover similarities between the iIoT platforms and to change them, in order to differentiate the platform from the competition.

The results of this paper are surely limited, concerning the subjectivity in the choice of search terms. Additional search terms could reveal additional relevant criteria, which were not identified. In addition, the structured literature search is limited by the exclusion of German articles from engineering conferences, which also have strong research interest in the integration of a platform approach for smart connected machine tools. The literature analysis also does not consider some viewing levels on platforms, as platform governance or interactions of the platform ecosystem with its environment [50, 51], which could also influence the criteria or extend the catalog. However, the elaborated criteria catalog could influence the platform governance, as it has a strong intersection with the degree of the openness and the architecture of a platform [10]. The limited search period providers additional constraints, as it does not consider the high dynamics in the market of iIoT platforms, excluding newest possible metrics or features, implemented and introduced during 2018. Therefore, criteria and corresponding solution-oriented metrics introduced after the end of the search period are not included in the catalog. Moreover, the application of the catalog still needs an empirical validation, which is not addressed in this paper. Against these limitations, a follow-up multiple-case study on market-ready iIoT platforms would provide potential topics for future research. Consequently, the application of the criteria catalogue and the evaluation of the application will be the next steps. These steps would provide additional managerial implications for the iIoT platform-providing companies and offer potential for further development of the catalog.

The results also provide a foundation for an empirical evaluation of the researched criteria. Manufacturing companies could be suitable candidates for an empirical study in the future with the goal to extend and to prioritize the researched criteria.

References

1. Westkämper, E., Löffler, C.: Strategien der Produktion. Technologien, Konzepte und Wege in die Praxis. Springer, Heidelberg (2016)
2. Roland Berger: Predictive Maintenance. Service der Zukunft – und wo er wirklich steht. Roland Berger (2017)
3. Staufen: Deutscher Industrie 4.0 Index 2017. Staufen Digital Neonex GmbH und Staugen AG (2017)
4. James, M., Chui, M., Bisson, P., Woetzel, J., Dobbs, R., Bughin, J., Aharon, D.: The Internet of Things: Mapping the value beyond the hype, McKinsey Global Institute (2015)
5. Porter, M. E., Heppelmann, J. E.: How smart, connected products are transforming competition. Harvard Business Review 92, 64–88 (2014)
6. Yoo, Y., Henfridsson, O., Lyytinen, K.: The new organizing logic of digital innovation: An agenda for information systems research. Information Systems Research 21(4), 364–380 (2010)
7. Lee, E., A.: Cyber physical Systems: Design challenges. In: Proceedings of the 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing. pp. 363-369. IEEE Computer Society, Washington (2008)
8. Schönhofen, F., Petrik, D., Schockert, S., Herzwurm, G.: Analysis of platform-based business models using quality function deployment. In: Proceedings of the 23rd International Symposium on Quality Function Deployment, pp. 65-80 (2017)
9. Krause, T., Strauß, O., Scheffler, G., Kett, H., Lehmann, K., Renner, T.: IT-Plattformen für das Internet der Dinge (IoT). Basis intelligenter Produkte und Services. Fraunhofer Verlag, Stuttgart (2017)
10. Tiwana, A.: Platform ecosystems: aligning architecture, governance, and strategy. Morgan Kaufman, Waltham (2013)
11. Gawer, A., Cusumano, M. A.: Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation. Harvard Business School Press, Boston (2002)
12. Bruhn, M., Hadwich, K.: Dienstleistungen 4.0: Erscheinungsformen, Transformationsprozesse und Managementimplikationen – Methoden – Instrumente. In: Bruhn, M., Handwich, K. (eds.) Dienstleistungen 4.0: Konzepte – Methoden – Instrumente, pp. 3-43. Springer Gabler, Wiesbaden (2017)
13. Gawer, A., Cusumano, M., A.: Industry Platforms and Ecosystem Innovation. Journal of Product Innovation Management 31(3), 417-433 (2014)
14. Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G:, Sihn, W., Ueda, K.: Cyber-physical systems in manufacturing. In: CIRP Annals – Manufacturing Technology 65, pp. 621-641 (2016)
15. Hartmann, M., Halecker, B.: Management of Innovation in the Industrial Internet of Things. In: Proceedings of the International Society for Professional Innovation Management (ISPIM), pp. 1-17 (2015)

16. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7), 1645-1660 (2013)
17. Serpanos, D., Wolf, M.: Internet-of-Things (IoT) Systems. Architectures, Algorithms, Methodologies. Springer Nature, Cham (2018)
18. Lenz, J., Wuest, T., Westkämper, E.: Holistic Approach to Machine Tool Data Analytics. Journal of Manufacturing Systems 48(C), 180-191 (2018)
19. Iansiti, M., Levien, R.: They Keystone Advanteage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability. Harvard Business Press, Boston (2004)
20. Rochet, J., Tirole, J.: Platform competition in two-sided markets. Journal of the European Economic Association 1(4), 990–1029 (2003)
21. Siemens, https://store.mindsphere.io/mindsphere/Applications, (Accessed: 06.12.2018)
22. PTC, https://www.ptc.com/en/marketplace, (Accessed: 06.12.2018)
23. Tranfield, D., Denyer, D., Smart, P.: Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management 14, 207–222 (2003)
24. Fink, A.: Conducting research literature reviews: from the internet to paper. Sage Publications, Los Angeles (2013)
25. Webster, J., Watson, R. T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26, XIII–XXIII (2002)
26. Rashman, L., Withers, E., Hartley, J.: Organizational learning and knowledge in public service organizations: A systematic review of the literature. International Journal of Management Reviews 11, 463–494 (2009)
27. Soni, G., Kodali, R.: A critical analysis of supply chain management content in empirical research. Business Process Management Journal 17, 238–266 (2011)
28. Winter, M., Knemeyer, A. M.: Exploring the integration of sustainability and supply chain management: Current state and opportunities for future inquiry. International Journal of Physical Distribution & Logistics Management 43, 18–38 (2013)
29. Axelsson, J., Papatheocharous, E., Andersson, J.: Characteristics of software ecosystems for Federated Embedded Systems: A case study. Information and Software Technology 56(11), 1457-1475 (2014)
30. Axelsson, J., Skoglund, M.: Quality assurance in software ecosystems: A systematic literature mapping and research agenda. Journal of Systems and Software 114, 68-81 (2016)
31. Babiceanu, R., F., Seker, R.: Big Data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. Computers in Industry 81, 128-137 (2016)
32. Chen, D.: A methodology for developing service in virtual manufacturing environment. Annual Reviews in Control 39, 102-117 (2015)
33. Eklund, V., Bosch, J.: Architecture for embedded open software ecosystems. Journal of Systems and Software 92, 143-148 (2014)
34. Eloranta, V., Orkoneva, L., Hakanen, E., Turunen, T.: Using platforms to pursue strategic opportunities in service-driven manufacturing. Service Science 8, 344-357 (2016)
35. Eloranta, V., Turunen, T.: Platforms in service-driven manufacturing: Leveraging complexity by connecting, sharing, and integration. Industrial Marketing Management 5, 178-186 (2016)
36. Farahzadi, A., Shams, P., Rezazadeh, J., Farahbakhsh, R.: Middleware Technologies for Cloud of Things – a survey. Digital Communications and Networks 4(3), 176-188 (2018)

37. Gawer, A.: Bridging differing perspectives on technological platforms: Toward an integrative framework. Research Policy 43, 1239-1249 (2014)
38. Herterich, M., M., Uebernickel, F., Brenner, W.: The impact of cyber-physical systems on industrial services in manufacturing. In: Procedia CIRP 30, pp. 323-328 (2015)
39. Hohwieler, E., Berger, R., Geisert, C.: Condition monitoring services for e-Maintenance. In: IFAC Proceedings Volumes 37, pp. 239-244 (2004)
40. Jesse, N.; Internet of Things and Big Data – The Disruption of the Value Chain and the Rise of New Software Ecosystems. In: IFAC-PapersOnLine 49(29), pp. 275-282 (2016)
41. Mikusz, M.: Towards a Conceptual Framework for Cyber-Physical Systems from the Service-Dominant Logic Perspective. In: Procedia CIRP 16, pp. 385-389 (2014)
42. Mikusz, M., Csiszar, A.: CPS Platform Approach to Industrial Robots: State of the Practice, Potentials, Future Research Directions. In: PACIS 2015 Proceedings, pp.176 (2015)
43. Shariazhadeh, N., Lundholm, T., Lindberg, L., Sivard, G.: Integration of digital factory with smart factory based on Internet of Things. In: Procedia CIRP 50, pp. 512-517 (2016)
44. Soeldner, C., Roth, A., Danziger, F., Moeslein, K.: Towards open innovation in embedded systems. In: Proceedings of the Nineteenth Americas Conference on Information Systems (AMCIS), (2013)
45. Tantik, E., Anderl, R.: Integrated Data Model and Structure for the Asset Administration Shell in Industrie 4.0. In: Procedia CIRP 60, pp. 86-91 (2017)
46. Tedeschi, S., Mehnen, J., Tapoglou, N., Rajkumar, R.: Security aspects in Cloud based condition monitoring of machine tools. In: Procedia CIRP 39, pp. 47-52 (2015)
47. Thames, L., Schaefer, D.: Software-defined cloud manufacturing for Industry 4.0. In: Procedia CIRP 52, pp. 12-17 (2016)
48. Waibel, M., W., Steenkamp, L., P., Moloko, N., Oosthuizen, G., A.: Investigating the effects of smart production systems on sustainability elements. In: Procedia Manufacturing 8, pp. 731-737 (2017)
49. Yue, X., Cai, H., Yan, H., Zou, C., Zhou, K.: Cloud-assisted industrial cyber-physical systems: an insight. Microprocessors and Microsystems 39(8), 109-118 (2005)
50. Mautsch, L., M.: Softwareplattformen für Unternehmenssoftwareökosysteme. Eul, Lohmar (2015)
51. Herzwurm, G.: Kundenorientierte Softwareproduktentwicklung. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden (2000)
52. De Reuver, M., Sorensen, M., Basole, R., C.: The digital platform:a research agenda. Journal of Information Technology 33(2), 124-135 (2018)

Special Track 03:

Demos & Prototyping

Track Chair:

Max Krüger
Universität Siegen

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Privacy-friendly User Location Tracking with Smart Devices: The BeaT Prototype

Jan H. Betzing[1], Marco Niemann[1], and C. Ingo Berendes[2]

[1] University of Münster, ERCIS, Münster, Germany
`{jan.betzing|marco.niemann}@ercis.de`
[2] Paderborn University, Paderborn, Germany
`Ingo.berendes@upb.de`

Abstract. Customers use smart devices to share their location data with service providers to co-create personalized, location-based services. However, mobile apps that record movement profiles not only yield value-added service but also bear potential for abuse. Especially apps utilizing GPS-based tracking pose a privacy risk because they could–once enabled–unnoticeably record data in private situations. In response, we developed a privacy-friendly solution, called *BeaT*, that tracks user locations without GPS and gives users full control over the time and scope of data collection. We leverage Bluetooth Beacon technology to confine the perimeter in which tracking takes place. This paper presents the requirements, algorithmic design, prototypical implementation, real-world use case, and evaluation setting for BeaT.

Keywords: User Location Tracking, Location-based Service, Privacy, Smart Device, Bluetooth Low Energy Beacon

Video: `https://www.smartmarketsquare.de/tracking-with-beat`

1 User Location Tracking–Friend or Foe?

Navigating by smart device or tracking a run in a fitness app are exemplary location-based services (LBS) where users share their location with service providers [1]. Even though service providers are limited in data exploration by data-protection regulations (e.g., required user consent and opt-in to data sharing), privacy risks and potentials for abuse remain [2,3]. Notably, once a mobile app has been granted permanent access to the GPS sensor, it can access the user's location even in private situations. As seen in the case of the fitness-tracking app Strava, unrestricted user tracking can–in the worst case–expose classified information such as the location of secret US military bases [4].

In response, this paper presents *BeaT, a privacy-friendly user location tracking mechanism for smart devices*, and its application in the smartmarket[2] research project, in which the authors developed a LBS for high street customers to track their shopping trips [see 3,5]. In addition to retail, BeaT can also be applied in other sectors such as healthcare or hospitality management to track time-logical user movement sequences.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Motivated by the disadvantages of existing solutions, we state the requirements for BeaT as follows: First, to prevent tracking in private situations, the solution must not use satellite or WiFi-based geolocation or access acoustic and visual sensors (R1). Additionally, the tracking perimeter should be defined ex-ante (R2). Second, while GPS only supports outdoor tracking, we want to track both indoor and outdoor locations (R3). Third, some LBS such as Foursquare use check-ins, i.e., users manually record the places they are visiting [1]. While this mechanism accounts for user control, it lacks convenience. Consequently, our solution should provide automatic tracking (R4) while retaining users' full control regarding the time and scope of data collection (R5). Fourth, users associate LBS such as navigation or AR games (e.g., Pokémon Go) with battery drain [6]. Our solution should limit mobile resource utilization regarding processing power, memory, (mobile) data, and battery to foster user acceptance (R6).

2 Prototypical Design and Implementation

2.1 Technology Selection and Setting

We selected Bluetooth Low Energy (BLE) beacons as the underlying tracking technology because (1) smart devices support the technology for a few years [7]; (2) beacons do not reach into private spaces (R1) as they have to be actively deployed in the perimeter within which tracking should commence (R2); (3) the BLE protocol has low mobile resource utilization (R6); and (4) beacons allow for both indoor and outdoor tracking (R3). Beacons are small low-range radio-frequency transmitters that frequently broadcast a predefined payload, consisting of a Universally Unique Identifier (UUID) and two alphanumeric values called major and minor [7]. Nearby devices can receive the signal, calculate their proximity to the beacon, and reason upon the payload [3].

In the smartmarket² case, customers can record their journey across multiple retailers that are geographically dispersed along the high street [5,8]. In return for their data, customers receive personalized offers from nearby stores. Beacon major values identify the store whereas beacon minor values indicate the purpose and location of a beacon (outside at an entrance, inside at a PoI, inside at a PoS) [3]. These purposes allow to distinguish customers passing by, entering a store, traversing a Point of Interest (PoI) such as a promotional display within a store, and visiting a Point of Sale (PoS). Additionally, a back-end database server holds the geocoordinates of each beacon.

Customers use a mobile app to record their shopping trips. On the first start, the app requests access to the Bluetooth sensor. Once granted, the device is able to listen continuously for beacon signals. However, smartmarket² is deployed individually for each high street to fulfill R2. Hence, the app only declares a single UUID to monitor. Given that 2^{122} UUID exist [7], it is unlikely to receive a matching beacon signal outside the predefined area. Concerning R4 and R5, smart devices can receive beacon signals automatically in the background, but customers have to actively start and stop tracking in the app before any signal is sensed. Additionally, the app lists active and past trips and allows to inspect and delete the collected information (R5). Lastly, current mobile OS indicate active Beacon tracking and inform users on continuous monitoring.

When the user starts tracking and relevant beacons are in range, the app receives a list of these beacons and their coarse distance (far, near, immediate) once every second. A naïve approach is to record this data entirely and transmit it to the back-end server. As this would pose substantial resource utilization and violate R6, we developed a solution to extract and send only relevant information from the raw data stream–BeaT.

2.2 The BEAcon Tracking (BeaT) Algorithm

Suppose we have an active trip `tr`, a set of recently seen majors `M` (representing stores), and a set of beacon sightings, henceforth called touches `T`. For each touch `t`, we call the procedure BEAT, given in Algorithm 1, to update customer journey information.

```
Input: (tr, M, t ∈ T)   ▷ tr is the current trip; M is a set of active majors; t is a newly received beacon touch
procedure BEAT
    te ← GETBYMAJOR(t.major)   ▷ Try to obtain a trip entry te based on the major of touch t
    if te != none and DIFF(te.end, t.time) < 10 min then
        tr.end ← t.time   ▷ Earmark current t as potential last t ∈ tr
        te.end ← t.time   ▷ Earmark current t as potential last t ∈ te
        switch t.type do
            case "entrance"
                if te.entrance_first == none then te.entrance_first ← t
                if t.dist == "immediate" then te.visited ← true   ▷ Mark te (store) as visited
                te.entrance_last ← t   ▷ Earmark t as potential last sighting
            case "PoI"
                if t.dist != "far" then
                    te.visited ← true   ▷ Mark te (store) as visited
                    if t ∉ te.poi then SENDTODB(t)   ▷ Send PoI touch t to the database server on first sight
                    te.poi ← te.poi ∪ t   ▷ Add t to the set te.poi of seen PoI
            case "PoS"
                if t.dist != "far" then te.visited ← true   ▷ Mark te (store) as visited
                if t.dist == "immediate" then
                    if te.pos_first == none then te.pos_first ← t
                    te.pos_last ← t   ▷ Earmark t as potential last sighting
                    if DIFF(te.pos_first, te.pos_last) > 30 sec then te.sale ← true
    else
        ENDEXPIREDTRIPENTRIES()   ▷ End expired te and remove respective te.major from M
        te ← CREATETRIPENTRY(t)   ▷ Create new trip entry te based on touch t
        M ← M ∪ t.major   ▷ Add t.major to active majors M
        tr.te ← tr.te ∪ te   ▷ Add te to current trip tr
        BEAT(t)   ▷ Execute BEAT again, having a te
```

Algorithm 1. BeaT

First we test, whether there was an interaction with this store (`t.major`) within the last ten minutes. If not, we create a new trip entry `te`, which represents one element in the customer journey, and add it to the trip `tr`. Also, inactive trip entries for stores that have not been seen in the last ten minutes are closed and their respective majors are removed from the set `M` of currently active majors. Depending on the type of touch (`t.type`) and the user's distance to the beacon (`t.dist`), the trip entry is updated. In effect, we store for each trip entry, when the store was first and last seen (`entrance_first|_last`), if the customer has entered the store (`visited`), a set of seen PoI, and if the customer has bought something (`sale`). The latter is approximated by testing, if the customer has been in immediate proximity to a PoS for at least 30 seconds. PoI sightings are directly transmitted to the back-end server so that it can potentially return relevant notifications for the user. Each trip is either terminated in-app by the user or by a timed function that closes a trip after three hours without beacon sightings. The trip `tr` and all associated `te` and `t` are transmitted to the back-end server en bloc and tracking is disabled.

3 Evaluation and Outlook

BeaT was added to the prototypical smartmarket² platform [5], which provides the socio-technical context to investigate user location tracking in the target high street environment. BeaT is the result of multiple cycles of building, testing and evaluation activities. We already asserted the general viability of beacon tracking in an artificial setting [3]. Advancing further along the continuum of Venable et al.'s *Human Risk & Effectiveness Strategy* [9] to evaluate IT artifacts, we will conduct a summative naturalistic evaluation in the fourth quarter of 2018. About 40 businesses in a German high street will install beacons and we invited about 400 high street customers to engage in customer journey tracking using BeaT over the course of two months. Besides judging how well BeaT achieves its envisioned environmental utility [9], we also designed our field test to shed light on the socio-technical issues of adoption and use of privacy-friendly LBS, which will subsequently yield additions to the knowledge base.

Large facilities such as hospitals, shopping malls, and airports already use BLE beacons to provide indoor navigation [7]. Besides its merits for high streets, BeaT could also constitute a viable extension to existing mobile service in these domains. For example, in hospitals, patients could use BeaT to record which doctors and wards (service providers) they have visited and–based on their current patient journey–receive further digital and personal service. Lastly, in future research, we will use the location data collected through BeaT to map, analyze, and predict customer journeys [8].

Acknowledgements

The authors thank the following undergraduate students from the University of Münster, who helped with designing and implementing the BeaT algorithm: Tim Hummelt, Leon Papke, David Pucka, Cedric Pumpe, Lennart Schäpermeier, and Pia Vollmer. This paper was developed in the research project smartmarket² (www.smartmarketsquare.de), which is funded by the German Federal Ministry of Education and Research (BMBF), promotion signs 02K15A073-02K15A074. The authors thank the Project Management Agency Karlsruhe (PTKA).

References

1. Dhar, S., Varshney, U.: Challenges and business models for mobile location-based services and advertising. Communications of the ACM 54(5), 121–129 (2011)
2. Schaub, F., Balebako, R., Durity, A.L., Cranor, L.F.: A Design Space for Effective Privacy Notices. In: Proceedings of the 11th Symposium on Usable Privacy and Security. pp. 1–17. SOUPS ’15 (2015)
3. Betzing, J.H.: Beacon-based Customer Tracking across the High Street: Perspectives for Location-based Smart Services in Retail. In: 24th Americas Conference on Information Systems. AMCIS ’18, New Orleans, LA, US (2018)
4. Woollacott, E.: Pentagon Bans GPS Fitness Apps, Says They’re ’Significant Risk’ To Troops. Forbes (aug 2018), https://www.forbes.com/sites/emmawoollacott/2018/08/07/pentagon-bans-gps-fitness-apps-says-theyre-significant-risk-to-troops/, last accessed: 2018-11-12
5. Bartelheimer, C., Betzing, J.H., Berendes, I., Beverungen, D.: Designing Multi-sided Community Platforms for Local High Street Retail. In: 26th European Conference on Information Systems. ECIS ’18, Portsmouth, UK (2018)
6. Faragher, R., Harle, R.: Location Fingerprinting with Bluetooth Low Energy Beacons. IEEE Journal on Selected Areas in Communications 33(11), 2418–2428 (2015)
7. Statler, S.: Beacon Technologies. Apress, Berkeley, CA (2016)
8. Berendes, I., Bartelheimer, C., Betzing, J.H., Beverungen, D.: Data-driven Customer Journey Mapping in Local High Streets: A Domain-specific Modeling Language. In: 39th International Conference on Information Systems. ICIS ’18, San Francisco, CA, USA (2018)
9. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: a Framework for Evaluation in Design Science Research. European Journal of Information Systems 25(1), 77–89 (2016)

Application-oriented robotics in nursing homes

Felix Carros[1], Rainer Wieching[1], Jens Lüssem[2], Lena Müller and Volker Wulf[1]

[1] Information Systems and New Media, University of Siegen, Siegen, Germany
{felix.carros,rainer.wieching,lena.mueller,volker.wulf}@uni-siegen.de
[2] Institute for Applied Computer Science, University of Applied Science Kiel, Kiel, Germany
{jens.luessem}@fh-kiel.de

Abstract. The ARiA project communicates about new working worlds in nursing with participatory approaches in order to establish innovative solution models for the challenges of demographic change in the field of robotics in nursing among the public and at the same time generate practice-relevant approaches for future products and services in this field. In the past, technical solutions of this kind were often developed and communicated in a primarily engineering logic without appropriate involvement of end users. Robot-based products and services and the associated working environments in nursing care for the elderly are changing rapidly. Potentials and barriers for future innovations in this field have to be made accessible and tangible to the specialist and to the public. The project ARiA aims to fulfill this with an open science approach.

Keywords: Robotics, elderly care, social robotics, human robot interaction, participatory design.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Augmented Reality for Set-up Processes

Sven Hoffmann[1], Aparecido Fabiano Pinatti de Carvalho[1]

[1] University of Siegen, Institute of Information Systems and New Media, Siegen, Germany
{sven.hoffmann,fabiano.pinatti}@uni-siegen.de

Abstract. This paper reports on the possibilities of an AR-based tool to support machine operators in a set-up scenario. The focus lies on the exchange and acquisition of knowledge-intensive human practices in the context of manual set-up processes on modern production machines. Thus we go further in comparison to well know AR use cases and implement AR-technology not only for visualisation purposes, but for knowledge exchange between workers.

Keywords: AR, Set-up, Design Case Study, Knowledge exchange

Mixed Reality for supporting Remote-Meetings

Florian Jasche[1], Jasmin Kirchhübel[2] & Thomas Ludwig[1]

[1] University of Siegen, Cyber-Physical Systems, Siegen, Germany
[2] University of Siegen, Information System and New Media, Siegen, Germany
{firstname.surname}@uni-siegen.de

Abstract. Nowadays meetings are no longer physically tied to one place. Especially in knowledge work, telephone calls or Skype conferences have long since complemented classic face-to-face meetings. Various research discourses, especially computer-supported group work, have been investigating for almost three decades how distributed group work can be supported in its various forms using IT. With the increasing performance of technologies focusing on Augmented Reality (AR) and Virtual Reality (VR), new possibilities have been added that offer a high potential for supporting distributed meetings. With this prototype, we present an approach that combines AR and VR to implement a communication system with various collaboration options for the appropriate support of distributed meetings. Our prototype focuses on scenarios in which two or more people are in the same room and one or more people are absent physically, but both parties can still cooperate remotely at the same time.

Keywords: Augmented Reality, Virtual Reality, Collaboration, Remote-Meetings

1 Introduction

Remote meetings have become a common feature of everyday work in many industrial and commercial contexts where cooperative tasks are required. Supporting such remote meetings through technology has long been one of the most important discourses within Computer-Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI) [1–6]. However, the problem of the appropriate degree of immersiveness has not gone away and representing work activities at a distance is still difficult. For instance, video conferences via Skype are based on face-to-face communications and lack support for dynamic and interactive activities such as work on whiteboards or with physical artifacts, which are important parts of practical collaboration during a meeting [7]. Other approaches focus on setting up complex structures equipped with multiple screens and cameras which offer an improved immersive experience, but are not necessarily applicable for mobile or ad hoc work tasks that are typical for modern cooperation [8–10].

With the increasing efficiency of technologies focusing on Augmented Reality (AR) and Virtual Reality (VR), new possibilities for supporting distributed meetings have opened up. Research to date already shows the possibilities of AR and VR for

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

improving remote cooperation [11–14]. However, the focus of current research remains for the most part either on VR as a digital meeting room or on AR as a supportive technology for face-to-face communication [15]. In our prototype we combine both technologies and contribute with insights into how to design an immersive approach that makes use of AR as well as VR to support remote cooperation within creative and object-oriented settings.

The system we present creates an immersive experience during remote meetings and offers additional collaboration options that are less fully realized in knwon video conferences. The goal is that the participants of the meeting, although they are at different locations physically, have the feeling of being together in a real room, using shared artifacts and being able to collaborate just like they would in face-to-face situations. Our prototype focuses on scenarios in which two or more people are in the same room and one or more people are absent but can cooperate remotely at the same time.

2 Description of the Prototype

In this section we present the implementation of the first iteration of our prototype. With regard to related research, we mainly focused on the topic of social presence and cooperation within remote meetings.

We built the entire system for the remote (VR) and local (AR) perspectives using the Unity game engine. For local participants, we used the Microsoft HoloLens, making use of its spatial mapping capability and hence its facilitation of free movement in the room, without being tied to a desk or immobile computer. For the remote person, we used the HTC Vive, because of its room scale tracking capability, which allows a more natural movement in the virtual environment. The remote person sees the virtual representation of an ideally well-known meeting room, which looks the same as the real room where the local participants meet. The virtual and real rooms are aligned by a visual marker, placed at a predefined position in the real room and scanned by each HoloLens user before joining the meeting session. Marker detection is done by Vuforia computer vision technology. The marker is used as a reference point to calculate the correct position for displaying the other participants and artifacts. All involved persons see a digital representation of the other participants as avatars exactly where they are in either the real or the virtual meeting room. Everyone is able to move freely and interact with each other naturally.

2.1 Creating a high sense of presence

One goal of our prototype is to involve (or at least create the illusion) the remote participants as actors in the exact same space as the local participants. We therefore try to satisfy the principal determinants of presence [16]. For further immersiveness, we created a fully virtual copy of the real room by using a combination of laser scanning and photogrammetry which, combined, delivers a highly detailed 3D model with almost photo-realistic textures. Besides visual information about the remote space, auditory cues are provided in the prototype. Therefore, the voice of all participants is transmitted as positional sound, in order to allow the users to locate people, even if they are currently not in their field of view.

We further implemented view independence for the remote person. The view independence is not limited to looking around, like in a 360-degree photo or video, because the VR-participant is also able to move independently in the virtual environment. The vividness of the environment is achieved through the presence of the other participants. Because the 3D model of the room is static, the environment itself is not interactive in its current state. With the goal to create a higher sense of presence, we therefore integrated specific elements in the room, so that all participants can interact with each other and the environment. Changes in these elements are transmitted to all other participants in real time. We assume that all these properties lead to a high degree of perceived presence, which should improve the involvement of the remote person.

2.2 Collaboration tools

As simple tools like a whiteboard help solving simple problems that are hard to articulate verbally, we implemented a virtual whiteboard. It is a simple white area in the size of classic whiteboards and it is placed at the same position where the real whiteboard is placed in the meeting room. The virtual whiteboard is visible to all participants, both remote and local, and everyone can draw on it. The drawing is transmitted in real time, so one can draw and explain his drawing at the same time. For our first iteration of the prototype we limited the functionality of the whiteboard to draw in four different predefined colors (black, red, blue and green), erase and to clear it completely. To select another color or to erase and clear, we placed six buttons on the left side of the board. The interface is looking the same locally as well as remotely, but the actual use differs through the interaction concepts of the hardware.

We also implemented a shared agenda for the current meeting. This agenda is placed on an empty wall of the room and is visible for all participants. The agenda is designed in a way that all participants are informed about the topics of the meeting. The agenda looks like a to-do-list where every user is able to check the tasks on the agenda.

An important factor of meetings is an appropriate time management. Schneider et al. [17] observed that the perception of time alters while being in a VR experience. To reduce this effect, we added an analog clock above the door of the meeting room, which is visible for all participants. Although our prototype has limited elements of this kind, we feel, that it carries some important implications with respect to the need to identify exactly what tools and implements need to be shared in what collaborative contexts. Our aim over time is to provide a library of virtual artifacts which can be used to construct a usable meeting room regardless of purpose.

2.3 Representation of the participants

In order to reach the goal of a more natural communication, all participants have to see each other or at least a representation of the participants. To keep the hardware setup simple, we chose three-dimensional avatars as a representation format. There is a wide range of different options for avatars. As Greenwald [18] pointed out, even very abstract avatars can enhance the perceived degree of social presence. We decided to use abstract avatars, which only consist of a head, a generic torso and hands. The head and hands are not connected to the torso. For the head we used two different abstract and low-poly 3D models of animal heads - a panda and a raccoon. Both models have clearly recognizable faces with eyes, but no mouth. Matching this, we implemented paws as hands. The pose of the HMDs is transmitted immediately to the head of the avatar. Even small motions like nodding are represented by the avatar. One is also able to recognize where someone is looking. This should lead to a more fluent and natural communication in the remote setting. This natural communication is supported by transmitting gestures. For the remote VR-participant, the poses of the controllers are transmitted to the paws of the avatar, but there is no animation for every single finger. On the local AR side, we use the hand tracking technology of the HoloLens. Unfortunately, the HoloLens only reports the position of the user's hand if it matches one of the predefined HoloLens gestures. The rotation of the paw is fixed and aligned to the yaw of the user's head because the HoloLens does not provide the rotation of the tracked hands. When the hands are not tracked by the HoloLens, the paws will disappear. The following figures give a short overview of the system:

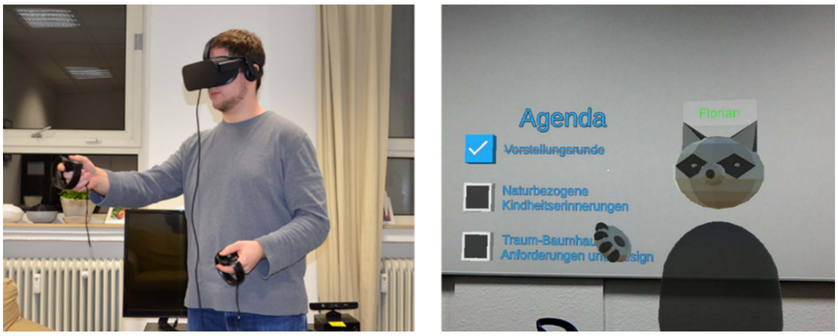


Figure 1: Left: Remote participant in his living room, right: his representation for the AR users.

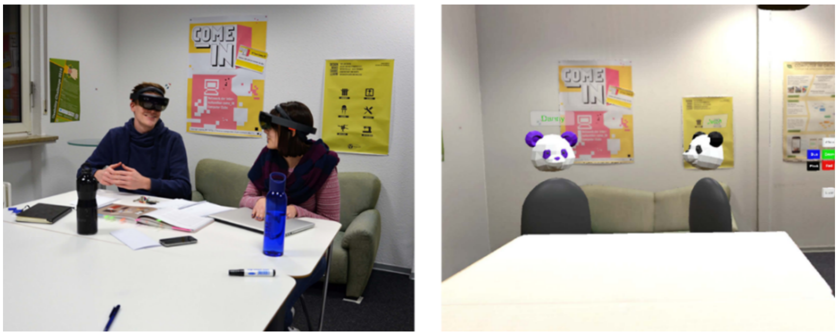


Figure 2: Left: Local participants in the meeting room, right: their representation for the VR users.

References

1. Gumienny, R., Gericke, L., Wenzel, M., Meinel, C.: Supporting creative collaboration in globally distributed companies. Proc. 2013 Conf. Comput. Support. Coop. Work - CSCW ’13. 995 (2013).
2. Koehne, B., Shih, P.C., Olson, J.S.: Remote and alone: Coping with being the remote member on the team. Proc. ACM Conf. Comput. Support. Coop. Work. CSCW. 1257–1266 (2012).
3. Lee, M.K., Takayama, L.: “Now, I Have a Body”: Uses and Social Norms for Mobile Remote Presence in the Workplace. In: Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11. p. 33. ACM Press, New York, New York, USA (2011).
4. Olson, J., Covi, L., Rocco, E., Miller, W., Allie, P.: A Room of Your Own: What would it take to help remote groups work as well as collocated groups? CHI 98 Cconference Summ. Hum. Factors Comput. Syst. 279–280 (1998).
5. Scholtz, J.C.: Telework: When your job is on the line. Conf. Hum. Factors Comput. Syst. - Proc. 22–27–Marc, 107–108 (1997).
6. Veinott, E.S., Olson, J., Olson, G.M., Fu, X.: Speakers Video Helps Remote Work : Who Need to Negotiate Common Ground from Seeing Each Other Benefit. Proc. SIGCHI Conf. Hum. Factors Comput. Syst. 302–309 (1999).
7. Hampton, S.E., Halpern, B.S., Winter, M., Balch, J.K., Parker, J.N., Baron, J.S., Palmer, M., Schildhauer, M.P., Bishop, P., Meagher, T.R., Specht, A.: Best Practices for Virtual Participation in Meetings: Experiences from Synthesis Centers. Bull. Ecol. Soc. Am. 98, 57–63 (2017).
8. Cisco: Cisco TelePresence IX5000 Series, https://www.cisco.com/c/en/us/products/collaboration-endpoints/ix5000-series/index.html.
9. Polycom: TELEPRESENCE WORKSPACE, http://www.polycom.com/your-workspace/immersive/telepresence.html.
10. Orts-Escolano, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P.A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S., Kang, S.B., Rhemann, C., Kohli, P., Lutchyn, Y., Keskin, C., Izadi, S., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S.: Holoportation: Virtual 3D Teleportation in Real-time. In: Proceedings of

the 29th Annual Symposium on User Interface Software and Technology - UIST ’16. pp. 741–754. ACM Press, New York, New York, USA (2016).
11. Gao, L., Bai, H., Lindeman, R., Billinghurst, M.: Static local environment capturing and sharing for MR remote collaboration. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA ’17. pp. 1–6 (2017).
12. Piumsomboon, T., Lee, G.A., Hart, J.D., Ens, B., Lindeman, R.W., Thomas, B.H., Billinghurst, M.: Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI ’18. pp. 1–13. ACM Press, New York, New York, USA (2018).
13. Schmalstieg, D., Szalavári, Z., Fuhrmann, A., Gervautz, M.: Studierstube: An environment for collaboration in augmented reality. Virtual Real. 3, 37–48 (1998).
14. Tait, M., Billinghurst, M.: The Effect of View Independence in a Collaborative AR System. Comput. Support. Coop. Work CSCW An Int. J. 24, 563–589 (2015).
15. Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., Billinghurst, M.: Exploring enhancements for remote mixed reality collaboration. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA ’17. pp. 1–5 (2017).
16. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence Teleoperators Virtual Environ. 7, 225–240 (1998).
17. Schneider, S.M., Kisby, C.K., Flint, E.P.: Effect of virtual reality on time perception in patients receiving chemotherapy. Support. Care Cancer. 19, 555–564 (2011).
18. Greenwald, S.W., Wang, Z., Funk, M., Maes, P.: Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. In: International Conference on Immersive Learning. pp. 75–90 (2017).

Gamification zur Motivationssteigerung von Werkern bei der Betriebsdatenerfassung

Christoph Kotthaus[1], Thomas Ludwig[1], Volkmar Pipek[1]

[1] Universität Siegen, Siegen, Deutschland
{christoph.kotthaus,thomas.ludwig,volkmar.pipek}@uni-siegen.de

Abstract. In Unternehmen bilden neben Stammdaten insbesondere Betriebsdaten die Basis für steuernde Maßnahmen bei der echtzeitnahen Produktionsplanung. Während Großunternehmen hierfür oft dedizierte Abteilungen und Technologien betreiben, ist dies für kleine und mittlere Unternehmen (KMU) oft unerschwinglich. Wenn überhaupt, werden Betriebsdaten nur handschriftlich erfasst. Die Weiterverarbeitung der Daten erfolgt allerdings meist zeitversetzt, sodass kurzfristige Handlungspotentiale ungenutzt bleiben. Außerdem stellt die Prozessdatenerfassung eine Nebentätigkeit für die Werker dar, wodurch die Dokumentation häufig entfällt, unvollständig oder fehlerhaft ist, zulasten der Datenqualität. Gegenstand dieser Arbeit ist, mittels Gamification die Motivation der Werker zu erhöhen, Betriebsdaten rechtzeitig und korrekt zu erfassen. Durch Interviews und Workshops mit Werkern und der Produktionsplanung eines KMU wurde ein praxisbasierter, digitaler Demonstrator entwickelt, worin Schichtmitarbeiter gemeinsam einen Avatar durch einen Parkour navigieren, indem alle ihre Betriebsdaten rechtzeitig und korrekt erfassen. Die Evaluation ergab ein ambivalentes Ergebnis, von zu spielerisch für ein als so ernst empfundenes Arbeitsfeld bis hin zu vielversprechend in Bezug auf die Betriebsdatenqualität.

Keywords: Gamification, HCI, Fallstudie, KMU, Usability

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Automatically Extracting and Analyzing Customer Needs from Twitter: A "Needmining" Prototype

Niklas Kühl[1], Dominik Martin[1], and Gerhard Satzger[1]

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Karlsruhe, Germany
{kuehl,martin,gerhard.satzger}@kit.edu

Abstract. Automated and scalable elicitation of customer needs is still in its infancy. With the proposed "Needmining" prototype, we aim to enable automated customer need extraction from the micro blog platform Twitter.

Keywords: Customer Needs, Twitter, Machine Learning, Need Elicitation

1 Introduction

For modern businesses to thrive, they must design customer-centric, innovative products and services [1]. To achieve this goal, providers first need to identify the needs of existing and potential customers—so they can directly address them within their marketing activities [2], and, more importantly, create innovative offerings which are tailored to those needs [3].

For *customer need elicitation*, the application of different methods like interviews, focus groups or conjoint analyses is typical [4]. These methods, however, can be very expensive and time-consuming [5]. This is mainly due to the manual effort required, as these traditional methods do neither scale well nor can they be automated.

Fortunately for businesses, customers are in most cases well aware of what they actually need [6]. Today, it is very common for customers to voluntarily and intrinsically share personal information via social media. Some of these social media instances may contain valuable insights about their concerns, wants, demands and problems [7]. These insights are publicly available and free of cost—but any manual review would be laborious and would only allow a snapshot at the time of the analysis.

In the proposed prototype, we show the feasibility to automatically identify and quantify customer needs from a publicly available source like Twitter. The developed artifact utilizes natural language processing and machine learning techniques to provide innovation managers with an aggregated overview of customer needs—and does so on an ongoing, "always current" basis. This contribution has the potential to drive a step change towards automated, analytical support for customer need elicitation.

2 The Needmining artifact

Within an overall Design Science Research (DSR) effort of automatically identifying customer needs from Twitter data [8], a first cycle evaluated the feasibility of "need tweet" identification as a basis. We implemented an automated identification

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

of tweets containing customer needs with supervised machine learning models for the exemplary domain of e-mobility [9]. The results from the technical experiment of this first cycle show the feasibility by providing superior statistical classification results compared to all regarded baselines [10]. Additionally, the best-working classification model is deployed as a single service. The second DSR cycle requires this service to output “need tweets” only, as we analyze possibilities on how to identify the needs themselves from this output. As a result of the second cycle, we achieve an additional service which is able to automatically assign tweets to multiple need categories for the domain of e-mobility [11]. Therefore, the previous two cycles demonstrate the feasibility of (a) automatic “need tweet” identification and (b) automatic quantification of the needs themselves. Apart from showing the statistical evaluations for each cycle, two isolated services are designed: One capable of identifying tweets containing needs, one capable of assigning these tweets to pre-defined need categories—based on supervised machine learning. Both services, however, are not tailored to a non-IT focused user, since all other activities would need to be undertaken manually.

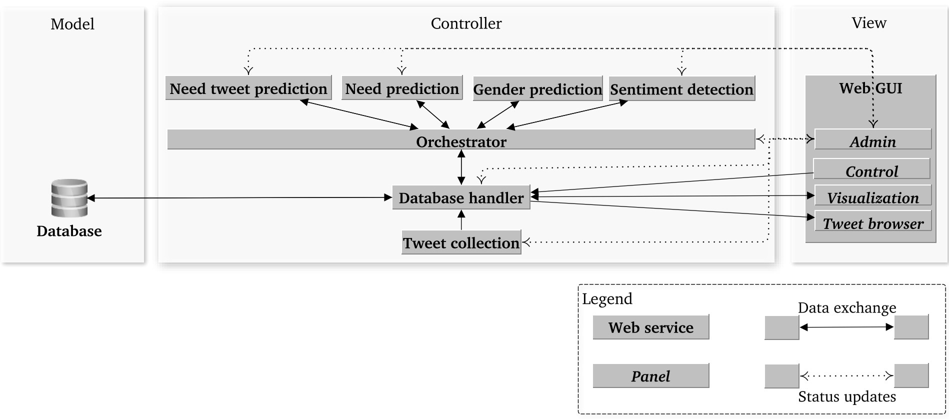


Figure 1. General architecture of the Needmining artifact

For the proposed prototype, we add a third DSR cycle to integrate both services into a comprehensive Needmining artifact. With this resulting artifact we later aim to evaluate not just statistical performance, but also (real) user acceptance. The Needmining artifact aims at automatically visualizing the most frequently expressed customer needs, which are autonomously posted on Twitter.

For a general architecture of the software design, we follow two principles: model-view-controller (MVC) [12] and service-oriented architecture (SOA) [13]. Figure 1 depicts the general overview of the architecture and its different web services. We structure our general architecture into three layers (model, view, controller) and a number of web services as independent instances, which can be accessed via a well-defined interface.

First, the tweets need to be received (*tweet collection*) and then stored for further processing. As multiple services need to access the storage, it is important to implement a *database handler*, which allows access to the database and monitors these accesses.

The *database* itself is part of the model layer. Apart from the tweet collection and the database handler, the controller layer consists of multiple other services. The *need tweet prediction* classifies whether a tweet from the database contains a need or not. In order to aggregate the needs themselves and achieve a distinct representation, they are automatically assigned to need categories as part of the *need prediction*. Moreover, the possibility of monitoring a change in needs over a specified period and differences due to demographic factors—like gender—can provide further insights into the development and the characteristics of certain needs. As an example, we additionally implement a self-built *gender prediction.* To automatically identify the sentiment of a tweet, we also utilize a *sentiment prediction* [14]. All these services (need tweet identification, need clustering, sentiment detection and gender detection) need to be orchestrated with regard to their access to databases as well as their sequence and error

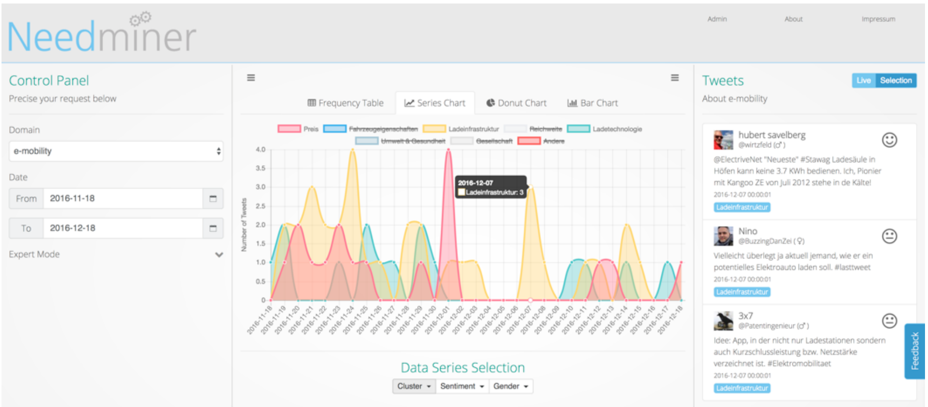


Figure 2. Screenshot from the Needminer GUI

handling. To address this, we implement a central *orchestrator* service, which orchestrates the main functions within the controller layer. As the orchestrator ensures a coherent processing of the data, the *Web GUI* from the view layer can directly access the data and visualize it for the end user (see figure 2). As the view layer is independent of the controller and the model, it can be enhanced or exchanged at any time, e.g. with a smartphone app interface.

3 Evaluation

As it is important to evaluate the artifact in a real-world scenario, we conduct an industry workshop with a large German utility and e-mobility services provider. The participants are from different divisions with different roles in the company. In total, seven employees take part in the three-hour workshop. The group consists of a product & innovation manager, a senior project manager, the head of a competence center, a group leader, a user researcher and an R&D project manager as well as an intern.

To evaluate the prototype, all participants can use the Needmining artifact for up to 30 minutes. Then, they fill out a further questionnaire regarding the experiences with using the software and we openly discuss positive and negative feedback. The questionnaire contains questions to assess (a) the helpfulness of the software, (b) its possible application within the company, (c) its ability to replace existing methods as well as (d) feedback regarding functionalities and design.

Out of seven participants, six regard the Needmining artifact as a helpful software. Furthermore, five participants can imagine using it for their daily business within their company. None of the participants states that the Needmining artifact in its current implementation can fully replace existing need elicitation methods. We further analyze the additional feedback, separated into functional and design feedback.

References

1. Teece, D.J.: Business Models, Business Strategy and Innovation. Long Range Plann. 43, 172 194 (2010).
2. Evans, J., Laskin, R.: The relationship marketing process: A conceptualization and application. Ind. Mark. Manag. 23, 439–452 (1994).
3. Slater, S.F.: Developing a customer value-based theory of the firm. J. Acad. Mark. Sci. 25, 162–167 (1997).
4. Hauser, J.R., Griffin, A.: The Voice of the Customer. Mark. Sci. 12, 1–27 (1993).
5. Blindheim, J., Wulvik, A., Steinert, M.: Using secondary video material for user observation in the needfinding process for new product development and design. In: Proceedings of the DESIGN 2016 14th International Design Conference (2016).
6. Mills, P.K., Margulies, N.: Toward a core typology of service organizations. Acad. Manag. Rev. 5, 255–266 (1980).
7. Misopoulos, F., Mitic, M., Kapoulas, A., Karapiperis, C.: Uncovering customer service experiences with Twitter: the case of airline industry. Manag. Decis. 52, 705–723 (2014).
8. Kuehl, N.: Needmining: Towards Analytical Support for Service Design. In: IESS 2016, Bucharest, Romania, May 25-27, 2016, Proceedings. pp. 187–200. Springer International Publishing, Cham (2016).
9. Scheurenbrand, J., Engel, C., Peters, F., Kühl, N.: Holistically Defining E-Mobility: A Modern Approach to Systematic Literature Reviews. In: 5th Karlsruhe Service Summit, Karlsruhe, Germany (2015).
10. Kühl, N., Scheurenbrand, J., Satzger, G.: Needmining: Identifying micro blog data containing customer needs. In: ECIS 2016 (2016).
11. Kühl, N., Mühlthaler, M., Goutier, M.: Automatically Quantifying Customer Need Tweets : Towards a Supervised Machine Learning Approach. In: HICSS-51, Waikoloa Village, Hawaii, United States, 3rd - 6th January 2018 (2018).
12. Kransner, G.E., Pope, S.: Cookbook for using the model-view-controller user interface paradigm. Object Oriented Program. 26–49 (1988).
13. Huhns, M.N., Singh, M.P.: Service-oriented computing: key concepts and principles. IEEE Internet Comput. 9, 75–81 (2005).
14. Thelwall, M.: Heart and soul: Sentiment strength detection in the social web with sentistrength. Proc. CyberEmotions. 5, 1–14 (2013).

GaNEsHA: Opportunities for Sustainable Transportation in Smart Cities

Johanna Meurer[1], Volker Wulf[1]

[1] University Siegen, Institute of Information Systems and New Media, Siegen, Germany
{first.author,second.author}@uni-siegen.de

Abstract. In developed nations, cities experience a growing pressure to establish sustainable transportation solutions. However, currently cities have still only little knowledge about the transportation routines, needs and troubles of its citizens. A greater awareness of daily transportation activities could support cities to build up customized transportation infrastructures that meet sustainable requirements. To address this gap, we build the GaNEsHA mobile application. A mobile crowed-sensing tool to track mobility activities. We used the tool as a probe in an interview study with 14 citizens. Our findings indicated a need for ICT design to move beyond normative tools that simply aim to 'encourage' people to adopt sustainable practices. Instead, we identified different transitions how municipal mobility services can support sustainable mobility practices. These include: (1) creating awareness of environmental footprints, (2) supporting alternative transportation options, (3) optimizing towards individual needs, (4) facilitating local communities and participative approaches.

Keywords: Sustainable mobility, mobile crowd-sensing; smart city, behavior change.

1 Background

Private transportation currently accounts for about a quarter of global CO_2 emissions. Given current trends, this value is set to increase by roughly 50% between now and the year 2030. In particular, many German cities are suffering from too high nitrogen values [1]. Throughout the western world, a (neo-) liberal agenda has arguably placed a greater burden upon individuals as ‘citizen-consumers’ to both consume for the sake of the market whilst accepting responsibility for the sustainability of their actions [4]. Over the last decade the sustainable HCI community has seen papers presenting interactive technologies that variously aim to support, inspire or persuade people to adopt pro-environmental behaviors [5]. A core assumption here is that the right kind of information about one’s behavior and its environmental effects will encourage more environmentally-friendly and sustainable habits [2]. However, measures aimed at changing individual behavior often face very low levels of uptake, complex constraints and even resistance [6]. Critique of these approaches often points to the normative

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

stance of assuming sustainable mobility is the only ‘good’ mobility, which risks being unintentionally paternalistic and behavioristic [3].

Addressing this issue, we are not questioning hot to better persuade citizen-consumers to conduct a more sustainable mobility, but instead to focus on the question how cities and its local municipal mobility services can provide appropriate transportation infrastructure that support multimodal sustainable mobility choices. Nowadays cities have only very little knowledge how its citizen experience sustainable mobility in their daily activities and how it can be optimally supported. Therefore, we question in this paper the notion of simply 'encouraging' people to adopt more sustainable mobility practices not from the normative perspective of individual citizens, but from the perspective of local municipal mobility services in cities.

2 Method and Prototype

To explore on the research question, we conducted interviews with 14 ‘ordinary’ citizens. Within these interviews we aimed to explore daily mobility practices while addressing interviewees “as analysists of their own and others’ practices”. To tackle this issue, we used a prototype to track mobility activities, then used the collected data as a probe to trigger detailed descriptions. The prototype provided an outline of users’ various journeys and associated modes of transport. We recorded the movements of all interviewees over a period of four weeks prior to the actual interviews taking place. This allowed us to review the rationales for various journeys in relation to concrete situations during the interviews. Additionally, the prototype offered some simple visualizations to sensitize interviewees to the possible environmental impact of their private mobility (see Figure 1). Overall, we found that the records helped interviewees to remember specific events, which served as prompts for stimulating reflections or to trigger richer accounts of particular journeys undertaken. Focusing on concrete situations also helped people to articulate specific rationales for the use of a particular mode of transport and the circumstances that led to it.

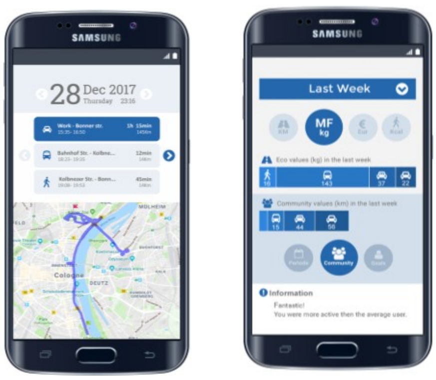


Figure 1. The picture on the left shows collected travel information. The picture on the right shows sustainability measurements. These designs were subsequently slightly modified to improve readability.

3 Results and Outlook

The analysis showed clearly that, although all participants spoke at some point during their interviews about the positive value of sustainability, sustainable mobility was not addressed as a driving factor in and for the organization of daily journeys. Moreover, we found that mobility practices were expressed to come with different practical exigencies that oriented to sustainable mobility in divergent ways. They incorporate: knowledge and information, self-reflection, thrift, comfort and collective responsibility. The prototype has helped to focus on these practical exigencies, how the interviewees actually handled sustainable mobility. We believe that the identification of practical exigencies is useful in order to use ICT to support sustainable mobility solutions beyond gross normative appeals (e.g. 'be sustainable!'). The points we have identified can be collected under more general themes that may inspire city services concerned with sustainable mobility. First of all, there is a need to design with a focus beyond a normative appeal of sustainable behavior, but to (1) creating awareness of environmental footprints. Secondly, we see opportunities by examining other values that are important for sustainable practices in (2) supporting alternative transportation options. Thirdly, this study offers a positive example to inspire the design regarding mobility practices and (3) optimizing towards infrastructural needs of the citizens. Lastly, we see a lot of potential in addressing different stakeholder groups who collaborate around processes of mutual learning and innovation (4) facilitating local communities & participative approaches. It would be interesting to study in future works more detailed how the prototype can be enhanced to take a deeper look at the daily mobility activities of the citizens. Cities could learn more about the concrete mobility needs and transport resources and how both can be better matched to support sustainable mobility.

References

1. Stewart Barr and Jan Prillwitz. 2012. Green travellers? Exploring the spatial context of sustainable mobility styles. Applied geography 32, 2: 798–809.
2. Eli Blevis. 2007. Sustainable interaction design: invention & disposal, renewal & reuse. In Proc. of the SIGCHI conf. on Human factors in Computing Systems, 503–512.
3. Hronn Brynjarsdottir, Maria Håkansson, James Pierce, Eric Baumer, Carl DiSalvo, and Phoebe Sengers. 2012. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In Proc. of the SIGCHI conf. on Human Factors in Computing Systems, 947–956.
4. John Clarke, Janet Newman, Nick Smith, Elizabeth Vidler, and Louise Westmarland. 2007. Creating citizen-consumers: Changing publics and changing public services. Pine Forge Press.
5. Paul Dourish. 2010. HCI and environmental sustainability: the politics of design and the design of politics. In Proceedings of the 8th ACM Conference on Designing Interactive Systems, 1–10.
6. James Pierce, Yolande Strengers, Phoebe Sengers, and Susanne Bødker. 2013. Introduction to the special issue on practice-oriented approaches to sustainable HCI. ACM Transactions on Computer-Human Interaction (TOCHI) 20, 4: 20.

TUCANA: A platform for using local processing power of edge devices for building data-driven services

Mirco Pyrtek[1], Sascha Xu[1], and Wolfgang Maass[1]

[1] German Research Center for Artificial Intelligence, 66123 Saarbrücken, Germany
{mirco.pyrtek,xiaguang_sascha.xu,wolfgang.maass}@dfki.de

Abstract. In the age of mobile cloud computing web-based systems are often designed to transfer data to large scaling online storage facilities in order to persistently save and analyze it with complex algorithms such as used in machine learning. These systems often require a reliable network connection, which does not hold for a variety of mobile business applications. As an alternative to traditional cloud-based systems the TUCANA approach makes use of the local processing power of mobile edge devices in order to come up with high complex AI pipelines processing data in real-time. By applying the idea of TUCANA to our service use-case called "nPotato" we developed an artificial, nociceptive potato that frequently measures and analyses acceleration data during the harvesting process of potatoes. In the given scenario sensory data is processed locally in real-time using the device's local computing power to gain higher productivity in the area of precision farming.

Keywords: Edge Computing, Internet of Things, Distributed AI, Web Technologies

1 Introduction

Mainstream mobile cloud computing architectures assume that local clients collect data and transfer this data to centralized data storage facilities. Major data processing is performed in these data storages and results are transferred to clients for presentation to users [1]. Examples are Google's Firebase or Microsoft's Azure.

These architectures normally assume high-speed Internet access so that data exchanges are performed without considerable latency. It also assumes that users are willing to transfer their data to centralized data centers, i.e. accept end-user license agreements (EULA) accordingly. These requirements do not always hold. In many areas, internet access is either unavailable or unreliable. This causes problems, for instance, for mobile business applications, such as used in farming.

With our new approach called "TUCANA" we aim at building a platform for the fast implementation of smart services that are working in closed local environments. This approach is based on modern approaches of mobile software development making use of distributed device capacities as given in the field of edge computing [2]. In this

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

scenario the execution of a service in a closed environment implies that a service does not rely on a stable internet connection and a high availability of computing capacities in cloud facilities since data measurement and data processing both can be done using the device's local sensors and processing power.

Moreover, since the communication of analysis results is important for most of the cases, the architecture comes with the opportunity to connect several local environments for sharing data through a secured peer-to-peer network . Each TUCANA environment (T-ENV) acts as a peer in this scenario. With the feature of connecting several T-ENVs there also comes the opportunity of building data processing pipelines making use of the capacities of multiple mobile edge devices and distributing the execution of AI models and other functionality within a connected network. With this approach computational complexity can be significantly reduced for single devices in their local environments but distributed to a variety of available peers in the network.

Applied to the context of real-time decision making in the area of farming such as described in [3] this means that problems based on low availability of computing resources during farming processes can be avoided in many cases. For instance, during the harvesting process of potatoes, connectivity errors of mobile monitoring devices harvested together with the crop such as the "nPotato" can occur because of a changing environment. If a measuring device used for monitoring the impact within the harvester needs to transfer large amounts of raw data for analysis to a central server, the connectivity most likely is not sufficient and the feedback to a farmer or a driver of the harvesting machine is either significantly delayed or not given at all. By processing the data directly on the measuring device, just the analysis results are necessarily transferred in order to give feedback to the farmer or the driver of the harvesting machine. This most likely results in a higher availability of the analysis results and provides the opportunity of reacting quickly to changing circumstances during the harvesting process.

2 Proposed demo

2.1 Demo description

Our demo showcases a cloud-independent AI platform for high complex problems, all solved in a protected, local software environment. Combining the sensory data and processing power of multiple mobile devices, TUCANA is a platform for distributed, privacy protecting smart services. Following this approach, we built a TUCANA application for the service use-case "nPotato". In this scenario the application assists a potato farmer during harvesting with real-time damage analysis and financial forecasting for market prices, all carried out on commonly available devices such as smartphones and laptops and without any necessary cloud infrastructure.

Figure 1. nPotato prototype

For realising this use-case we use a standard smartphone as a measuring device in and run our TUCANA application in it's Google Chrome Web Browser. After the smartphone is set up and running we put it into a case built by a 3D printer as shown in figure 1. In order to visualize what is going on inside the potato case we use the same web application on a tablet PC's Chrome Browser that is connected to the nPotato smartphone as a peer. Data sensed and analysed by the smartphone inside the case gets transferred through a peer-to-peer network to the remote tablet by using a technology called WebRTC[1]. The user of the tablet PC, in our case the driver of a potato harvesting machine or the farmer himself, now has access to realtime information about the current quality of the crop as well as the total expected yield. These predictions are based on the analyzed sensory data of the nPotato phone and directly sent to the tablet PC. Additionally, the data obtained from several connected potatoes is processed further on the tablet by analysing it in combination with future potato price estimates in order to give a prediction about upcoming incomes for the farmer.

Figure 2. Reliability testing during potato harvesting

Furthermore, we already tested the reliability of the application in a real world setting by harvesting the artificial potato with a real potato harvesting machine as shown in figure 2. This picture was shot during potato harvesting at a local farmer. To show the system also in controlled environments we developed a demo case. The demo proceeds as follows:

[1] see https://webrtc.org/ for further details

1. The smartphone is inserted into the case and connected to the remote tablet PC.
2. The application is started remotely through the tablet PC.
3. The nPotato is dropped from several heights and communicates the analysis results to the tablet PC.
4. The user gets information about the status of the potato in real-time.
5. The data is further analyzed on the tablet and the estimated monetary income for a farmer is shown.

2.2 Requirements

The demo has only minor technical requirements. It needs a space for executing the fall experiments with the nPotato. We usually used a tapeline for measuring the fall height and a pillow for reducing the noise of the fall experiments. Additionally the devices need access to a wifi access point as a basis for the secured peer-to-peer communication. Optionally, a larger screen for our tablet PC could improve the demonstration for the visitors.

2.3 Interaction Possibilities

Visitors can interact with our demo system in several ways. We provide nPotato cases with smartphones inside, so visitors can test the damage and fall detection on their own. Since our application is built as a web application, no installation is required to use it. If someone is interested in connecting to our system either as a potato measuring device or as the remote device that gets analysis results from several potatoes he or she can either use his or her own smartphone (Android required) or one of our phones we use to demonstrate the technology to interact with the system and process and transfer data on the device.

References

1. Dinh, H. T., Lee, C., Niyato, D., & Wang, P. (2013). A survey of mobile cloud computing: architecture, applications, and approaches. Wireless communications and mobile computing, 13(18), pp. 1587-1611, (2013).
2. Hu, Y. C.; Patel, M.; Sabella, D.; Sprecher, N.; Young, V.: Mobile edge computing. A key technology towards 5G. ETSI White Paper, 11 (11), pp. 1 - 16, (2015)
3. Maaß, W., Shcherbatyi, I., Marquardt, S.: Real-time Decision Making with Smart Farming Services. Proc. of 75th International Conference on Agricultural Engineering (AgEng), (2017)

Demonstrator zur Beschreibung und Visualisierung einer kritischen Infrastruktur

André Sekulla[1], Christopher Schmitz[2], Sebastian Pape[2], and Volkmar Pipek[1]

[1] University of Siegen, Institute of Information Systems, Germany
{andre.sekulla,volkmar.pipek}@uni-siegen.de
[2] Goethe University Frankfurt, Chair of Mobile Business & Multilateral Security, Germany
{christopher.schmitz,sebastian.pape}@m-chair.de

Abstract. Kleine und mittelständische Unternehmen haben oftmals große Probleme in der Einführung eines ISMS und einer Verbesserung der eigenen IT-Sicherheit. Besonders kritische Infrastrukturen wie Energienetzbetreiber sind gesetzlich dazu verpflichtet ein ISMS einzuführen und für eine bestmögliche Sicherheit zu sorgen. Diesen Unternehmen fehlt es jedoch oftmals an möglichen Mitteln und Erfahrungen. Mit dem Forschungsprojekt SIDATE soll dieses Wissensdefizit untersucht und analysiert werden. Es wurden mehrere leichtgewichtige Werkzeuge entwickelt, die auf einer zentralen Wissens- und Austauschplattform angesiedelt sind. So sollen mit der Entwicklung einer leichtgewichtigen Beschreibungssprache und dem zugehörigen Demonstrator kleine und mittelständische Energienetzbetreiber unterstützt werden. Mit dieser Arbeit möchten wir den Demonstrator zur Beschreibung, Visualisierung und Untersuchung auf IT-Sicherheit einer kritischen Infrastruktur vorstellen. Mit der Beschreibungssprache ist es zudem möglich Angriffszenarien und Angriffsbäume zu visualisieren, um mögliche Sicherheitslücken besser aufdecken zu können. Insbesondere soll dieses Werkzeug durch Exportfunktionen einen Erfahrungsaustausch zwischen den Nutzern bzw. Energienetzbetreibern fördern und eine Steigerung der IT-Sicherheit stärken.

Keywords: IT-Sicherheit, Kritische Infrastruktur, Visuelle Beschreibungssprache

Acknowledgements

Vielen Dank an Leon Rische und Florian Becker für die Zusammenarbeit bei der technischen Umsetzung. Dieser Demonstrator wurde im Rahmen des Projekts SIDATE entwickelt, das vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Förderschwerpunktes "IT-Sicherheit für kritische Infrastrukturen" gefördert wird. Förderungsnummer: 16KIS0239K, 16KIS0240.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Entwicklung einer alltagsnahen persuasiven App zur Bewegungsmotivation für ältere Nutzerinnen und Nutzer

David Struzek[1], Claudia Müller[1] and Alexander Boden[2]

[1] University of Siegen, Information Systems and New Media, 57072 Siegen, Germany
{david.struzek,claudia.mueller}@uni-siegen.de
[2] Fraunhofer Institute for Applied Information Technology FIT,
Schloss Birlinghoven, St. Augustin, Germany
alexander.boden@fit.fraunhofer.de

Abstract. This paper intends to give a short overview on the development of a persuasive widget system to increase the level of physical activity in the context of participatory IT research for and with older adults. The complete work was embedded in the three-year research project Cognitive Village.

Keywords: Alternde Gesellschaft, Persuasive Systeme, Gesundheit, Sensorbasierte Systeme, Bewegungsmotivation, Participatory Design Approach

1 Einführung

Ältere Menschen prägen zunehmend das Gesellschaftsbild und damit auch den Alltag aller Bürger. Demnach ist jeder vierte Deutsche über 60 Jahre, jeder fünfte über 65, was im Jahre 2014 einen Anteil von 21% ausmachte [1]. Industrie wie auch Forschung versuchen dem entgegenzukommen und mit moderner Technik, z.B. Ambient Assisted Living (AAL) Systemen, ältere oder pflegebedürftige Menschen weitgehend zu unterstützen. Anwendungen im Bereich der Prävention, u.a. in den Feldern Ernährung und Bewegung, erhalten zunehmend Aufmerksamkeit als Beitrag für ein längeres und beschwerdefreies Leben [2]. Adaptive persuasive Systeme stehen dabei im Fokus, die Inhalte, die auf der Grundlage persuasiver Kommunikation beruhen, an das Verhalten, die Erfahrungen oder die Umgebung der Nutzer anpassen und damit Verhaltensänderungen stimulieren [3]. Die Arbeit beschreibt ein sensor-basiertes Widget in Form einer Blume, das zur Bewegungssteigerung für Senioren innerhalb des dreijährigen BMBF- geförderten Forschungsprojektes "Cognitive Village“ [4] auf Grundlage des Participatory Design Ansatzes entwickelt und implementiert wurde.

2 Stand der Forschung

Es fällt deutlich auf, dass viele AAL-Projekte im Feld der persuasiven Systeme nur wenig (qualitativ-) empirisch vorgehen, und die Konzipierung- und Gestaltungsarbeit somit eher auf bestehende Literatur oder vorliegende Statistiken stützen [5]. Die

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Herangehensweise ist nachvollziehbar, jedoch kann die oft fehlende Praxis bemängelt werden, da viele Konzepte erst im realen Alltagskontext ihre Wirksamkeit unter Beweis stellen. Daher fordern Autoren eine intensive Einbeziehung der Stakeholder in den Designprozess, beispielsweise mit Methoden des Participatory Design [6], [7]. Arbeiten wie von de Oliveira et al. (2010) [8] und Kumahara und Mori (2014) [9] zeigen, wie weit viele Projekte in der Entwicklung von persuasiven Systemen von den potentiellen Nutzern und realen Nutzungskontexten entfernt sind. Damit die Gestaltung persuasiver Systeme in Praxis funktioniert und nachhaltigen Erfolg hat, ist es besonders wichtig, konkrete Bedürfnisse gemeinsam mit den NutzerInnen zu verstehen, zu co-explorieren und Motivationsstrategien zur Verhaltensänderung entsprechen behutsam in Alltagspraktiken und –vorstellungen einzubetten [10].

3 Anforderungsanalyse

Zur Ermittlung der Nutzungsanforderungen wurde mit zehn TeilnehmerInnen im Alter von 67 bis 82 Jahren gearbeitet, die im Laufe des Forschungsprojekts unterschiedliche digitale Geräte erhalten haben. Dazu wurden verschiedene qualitative Methoden wie Fokusgruppensitzungen, als auch Interviews und Technology Probes [11] eingesetzt. Aus den Bedürfnissen konnten im weiteren Verlauf an das persuasive System Anforderungen abgeleitet und priorisiert werden. Dazu gehörten u.a. die Darstellung und Personalisierung des Systems, das Autonomiebewusstsein, die Kennzahlen z.B. aktuelle Schritte, die Abwesenheit von Ermahnungen und geringe Nutzungshürden.

4 Designidee und Funktionsbeschreibung

Auf Basis der Anforderungen wurden vom Autor potentielle Designideen konzipiert, eingegrenzt und mit Funktionen erweitert. Die ausgewählte Idee Blumenwidget bestand aus neun Elementen, die bis zum finalen Prototypen in abgeänderter Form übernommen wurden.

Fensterscheibe. Am Fenster wird immer das aktuelle lokale Wetter anhand der Postleitzahl angezeigt, wodurch sich auch die Fensterscheibe verändert. Für besondere Tage, wie dem Geburtstag des Nutzers oder an Weihnachten, werden besondere passende Elemente im Hintergrund oder auf der Fensterbank platziert, um die Gesamtszene personalisierter zu gestalten.

Topfpflanze. Die Topfpflanze ist das primäre Element des Widgets. Die Blume wächst in der Anfangsphase in 20% Schritten und symbolisiert immer die am letzten Tag gelaufenen Schritte. Dabei hängt die Wachstumsgröße von dem Füllzustand der Wasserkanne ab. Es gibt verschiedene Blumen und Farben. Bei 100% soll die Blume verhältnismäßig groß und kraftvoll wirken, damit der Nutzer ein positives Nutzererlebnis verspürt und erfreut ist, den Tag zuvor sein Tagesziel erreicht zu haben.

Wasserkanne. Die Wasserkanne zeigt die aktuelle Schrittzahl in 20% Etappen an, die zuvor individuell eingestellt wurde. Je mehr Wasser mit den gegangenen Schritten gesammelt wurde, umso stärker blüht die Blume am nächsten Tag auf. Morgens wird die Kanne geleert. Zusätzlich soll die Wasserkanne den Nutzer zum Trinken animieren und in einfacher Weise aufzeigen wie viel der Nutzer sich bewegt hat.

Post-Ist/Sticker (Achievements). Post-Ist oder Sticker, sind kleine aufmunternde Nachrichten für den Nutzer und zeigen bildlich erreichte Tagesziele der gesamten Woche auf. Diese kleben im Hintergrund an der Fensterscheibe. Da es mehrere Stickervarianten gibt und diese in zufälliger Reihenfolge erscheinen, sehen gesammelte Sticker im Gesamtbild bei jedem Nutzer anders aus. Zu Beginn einer neuen Woche verschwinden alte oder bisher gesammelte Sticker und das Sammeln fängt von neuem an. Hat der Nutzer dessen ungeachtet drei Sticker gesammelt, hat er die Möglichkeit Pokale zu erwerben.

Kalender. Der Kalender zeigt dem Nutzer das aktuelle Datum an. Dabei geht es vor allem um den letzten synchronisierten Stand.

Notizbuch. Das Notizbuch zeigt die aktuelle und auf Wunsch auch die persönlichen Maximalschritte an.

Einstellungen. In den Einstellungen kann der Nutzer seinen Namen, seine Zielschritte und seine Postleitzahl eintragen, damit das lokale Wetter abgerufen werden kann. Zudem kann der Nutzer auch eine Auswahl an Blumen vorfinden, für die er sich entscheiden kann.

Titel. In den Einstellungen kann der Nutzer dem Widget einen persönlichen Titel vergeben, wenn er damit einverstanden ist. Damit sollte das Widget noch persönlicher wirken, als ein generischer Widget-Name.

Verschiedene Modi. Neben den geplanten interaktiven Elementen werden auch verschiedene Modi eingesetzt, die den Inhalt des Widgets realitätsnaher aussehen lassen. Dadurch soll die Nutzungsakzeptanz des Widgets erhöht und die kritische Distanz reduziert werden, was bei älteren Menschen als Zielgruppe nicht außer Acht gelassen werden darf. Demgemäß wird zwischen fünf verschiedenen Modi unterschieden, zwischen denen das Widget je nach Ort, Uhrzeit und Datum wechselt.

4.1 Implementierung der Hardware und Softwarearchitektur

Für die Nutzung werden die Sensoren einer Smartwatch (Huawei Watch [12] und Nokia Go [13]) verwendet und damit Daten (Schritte) generiert, die am Ende eines iterativen Datenverarbeitungsprozesses in Form einer Veränderung des Blumen-Widgets, das innerhalb des Cognitive Village Dashboards open.Dash [14] eingebunden wurde, betrachtet wird. Open.Dash ist ein Open Source Visualisierungsrahmenwerk, das hauptsächlich in Javascript geschrieben wurde. Es bietet Nutzern ein Dashboard, das an bereits existierende Datenquellen angeschlossen ist, eine Nutzerverwaltung und andere http/websocket Quellen. Die Nutzung des Dashboards samt Widgets erfolgte über einen Webbrowser, im Falle von Cognitive Village über ein Samsung Tablet. Bewegt sich der Nutzer, so dass die Uhr eine Veränderung feststellt, werden über einen automatisierten Prozess die generierten Nutzerdaten an eine Middleware gesendet. Zur Synchronisation, Speicherung und Weiterleitung der Daten, wurde ein Smartphone eingesetzt. Die Middleware empfängt die Daten, verarbeitet sie und teilt sie den jeweiligen Funktionen zu. Dabei synchronisieren die Geräte alle 10 Sekunden miteinander und speichern die Daten, auch bei Internetausfall, auf dem Smartphone.

5 Zusammenfassung

Alle einzelnen Designideen und –iterationen wurden in enger Zusammenarbeit mit der Seniorengruppe exploriert, entwickelt und schließlich über mehrere Wochen genutzt. Der starke partizipative und alltagsnahe Ansatz, der sich auch in der Wahl der visuellen Elemente zeigt, hat in erfolgreicher Weise zu einer Annahme und längerfristigen Nutzungsmotivation geführt.

Referenzen

1. Statistisches Bundesamt: Ältere Menschen in Deutschland und der EU. (2016). – https://www.destatis.de/. Sichtung: 20.06.2017
2. Statistik Austria: Vorausberechnete Bevölkerungsstruktur für Österreich 2015-2100 laut Hauptszenario. (2016). – http://www.statistik.at/ web_de/statistiken/bevoelkerung/demographische_prognosen/ bevoelkerungsprognosen/027308.html. Sichtung: 07.03.2017
3. Enzyklo.de: adaptives System. (2014). http://www.enzyklo.de/ Begriff/adaptivesSystem. – Sichtung: 25.08.2017
4. cognitive-village.de, Forschungskolleg Universität Siegen: Cognitive Village - Vernetztes Dorf. (2016). http://www.cognitive-village.de/. – Sichtung: 20.06.2017
5. ALL-Deutschland.de: Was ist Ambient Assisted Living? (2016). http:// www.aal-deutschland.de/aal-1. – Sichtung: 20.06.2017
6. Hirschheim, R.; Klein, H. K.: Realizing Emancipatory Principles in Information Systems Development: The Case for ETHICS. 18. 1994
7. Chatterjee, S.; Price, A.: Healthy Living with Persuasive Technologies: Framework, Issues, and Challenges. In: Journal of the American Medical Informatics Association JAMIA 16 (2009), Nr. 2, 171-178. http://dx.doi.org/doi: 10.1197/jamia.M2859.. – DOI doi:10.1197/jamia.M2859.
8. Kaptein, Maurits; De Ruyter, Boris; Markopoulos, Panos; Aarts, Emile: Adaptive Persuasive Systems: A Study of Tailored Persuasive Text Messages to Reduce Snacking. In: ACM Transactions on Interactive Intelligent Systems 2 (2012), Nr. 2, Article 10. http://dx.doi.org/doi:10.1145/2209310.2209313. – DOI doi:10.1145/2209310.2209313
9. Consolvo, S; Markle, K.; Patrick, K. K. C. K. Chanasyk: Designing for persuasion: Mobile services for health behavior change. (2009). – In Proceedings of the 4th International Conference on Persuasive Technology (PERSUASIVE).

10. Patrick, K.; Raab, F.; Adams, M. A.; Dillon, L.; Zabinski, M.; Rock, C. L.; Griswold, W. G.; Norman, G. J.: A text message based intervention for weight loss: Randomized controlled trial. In: J. Med. Internet Res. 11 (2009), S. 1–9
11. M ller, Claudia; Schorch, Mar n; Struzek, David; Neumann, Marleen: Tech- nology Probes als Mittel zur Unterst tzung der Technik-Aneignung. In: Burg- hardt, Manuel (Hrsg.); Wimmer, Raphael (Hrsg.); Wolff, Christian (Hrsg.); Womser-Hacker, Christa (Hrsg.): Mensch und Computer 2017 - Workshopband. Gesellschaft f r Informatik e.V., Regensburg, 2017
12. HUAWEI WATCH: HUAWEI WATCH. http://consumer.huawei.com/ de/wearables/watch/. Version: 2017. – Sichtung: 22.10.2017
13. Nokia Go: Nokia Go - Aktivitäts- und Schlaftracker. https://health.nokia. com/de/de/go. Version: 2017. – Sichtung: 21.10.2017
14. open.Dash: Vordefinierte Visualisierungselemente zum einfachen Erstellen von Smart Home Ansichten und der Möglichkeiten zur Exploration von Daten und freien Gestaltung individueller Visualisierungselemente. (2016). https:// opendash.de/. – Sichtung: 20.06.2017

A browser-based modeling tool for studying the learning of conceptual modeling based on a multi-modal data collection approach

Benjamin Ternes[1], Stefan Strecker[1], Kristina Rosenthal[1], and Hagen Barth[1]

[1]University of Hagen, Enterprise Modelling Research Group, Hagen, Germany
{benjamin.ternes,stefan.strecker,kristina.rosenthal,
hagen.barth}@fernuni-hagen.de

Abstract. How do we learn conceptual modeling? What are common learning difficulties? Which tool support assists learners in what respect? The paper at hand reports on the design and development of a browser-based modeling tool integrated with a learning observatory in support of learning conceptual modeling and of studying the learning of conceptual modeling. Implementing a multi-modal data collection approach, the learning observatory tracks learner-tool interactions, records verbal data from learners and surveys learners about their learning processes to provide for analyses at the individual and aggregate learner levels in the quest for identifying patterns of learning processes, learning barriers and difficulties. We report on the current state of prototype development, discuss its software architecture and outline future development steps.

Keywords: Conceptual modeling, Modeling tool, Modeling tool development, Prototyping.

1 Introduction

Viewed as an activity, conceptual modeling involves an intricate array of cognitive processes and performed actions including abstracting, conceptualizing, associating, contextualizing, interpreting & sense-making, judging & evaluating, drawing & visualizing, and, in group settings, communicating, discussing and agreeing. Learning conceptual modeling is, hence, a complex task for learners associated with codified as well as tacit knowledge and a learning process involving knowledge acquisition through experience (e.g., [1]). Learning conceptual modeling involves mastering theoretical foundations, modeling methods and languages, and applying them to modeling tasks by critically reflecting on domain-specific technical language in the light of set modeling objectives [2, 3].

Research on learning conceptual modeling, hence, seems likely to benefit from taking multiple complementary angles on studying learning processes to account for their multifaceted nature. Related work on learning support tools include process-oriented feedback [4, 5], gamification [6] and serious games [7] as well as work on the collection, aggregation, analysis and evaluation of data on learners and their learning context

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

subsumed under the term learning analytics [8, 9]. Complementing prior work on tool development for supporting and studying learning conceptual modeling, we initiated a long-term research program to better understand how novice modelers learn a modeling language respectively modeling method and how tool support assists learners in what respect. The research program is based on the fundamental assumption that learning processes of novice modelers deserve study from several complementary perspectives. Therefore, it bases on mixed method research designs involving multi-modal data collection.

As part of that research program, we have been developing a learning observatory integrated with a modeling tool in support of learning conceptual modeling and of studying the learning of conceptual modeling. Overarching research objective guiding the implementation efforts is to identify patterns of learning processes, common learning barriers and learning difficulties. For that purpose, the learning observatory currently supports three modes of data collection on learners' learning processes: (1) tracking and recording (every) learner-tool interaction; (2) conducting written pre- and post-modeling online surveys of learners; (3) recording verbal data from learners while modeling (supporting 'think-aloud'-like data generation methods [10]).

The primary application context of the modeling tool and learning observatory are two introductory courses on conceptual modeling in a distance learning and teaching setting with large cohorts of students (1000+ and 200+ per semester). Based on a blended learning approach, these two courses presently pick out data modeling (in a variant of the Entity-Relationship Model), object-oriented modeling (a tailored subset of UML Class Diagrams) and business process modeling (a subset of the MEMO OrgML [11]). Development of the modeling tool is, consequently, tailored to the respective taught modeling languages and requirements for software development following from this specific application setting.

2 Prototype Presentation

The current working prototype has been under development since 2013 to explore design and implementation strategies for tracking interactions of learners with graphical editors. Two essential requirements drive the prototype development: (i) Platform independence as well as (ii) usability (especially an intuitive user interface). Technically, we opted for a web application with a JavaScript-driven browser frontend and a Java EE (Enterprise Edition)-based backend (see [12, 13]). The frontend design considerations, operating principles and essential requirements are outlined in [13]. A revised and extended version of the software architecture is depicted in Fig. 1. Note that the frontend user interface currently implements two graphical editors: First, an editor for constructing ER diagrams (in the variant used in the two introductory courses) and a second editor for creating MEMO control-flow and decomposition diagrams [11].

As a first step to realize tool support for learning conceptual modeling in the introductory courses, we have implemented an ad-hoc syntax validation to verify the syntactic correctness of conceptual models. Feedback on syntax errors is provided via natural language comments and the highlighting of graphical notation symbols. Additionally, a video-based step-by-step tutorial comprising short video excerpts has been implemented aimed at accelerating and fostering the general handling of the tool. Moreover, the tutorial provides an open discussion forum for learners to discuss emerging difficulties, e.g., with other learners while working on modeling tasks or to receive technical support on issues and news concerning the tool.

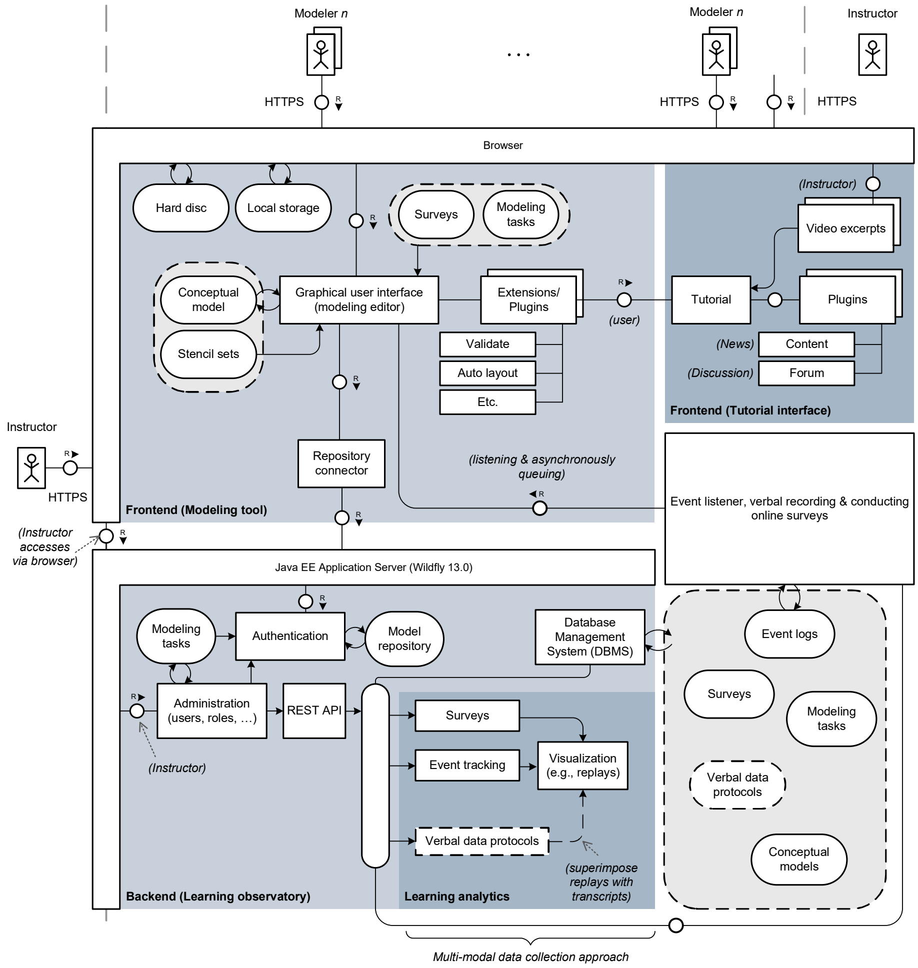


Figure 1. Revised software architecture at the conceptual design level depicted using FMC (Fundamental Modeling Concepts, see [14])

To include several complementary perspectives on learners' learning processes, we extended the learning observatory with further data collection approaches. Particularly, the learning analytics component has been extended for tracking, analyzing and evaluating learner-editor interactions, e.g., algorithms for tracking events. More specifically, the prototype implements an algorithm which tracks the learner-editor interactions, while working on, e.g., modeling tasks, by storing every learner-tool interaction with the graphical editor as a timed discrete event (*modelid, modelElementid, timestamp, userid, operation, type, content*). The tracking data is subsequently used for the learning analytics component. Currently, the learning analytics analysis and visualization component supports a step-by-step replay (i.e., visually showing every step of model construction), an automatic replay (i.e., visually showing model construction in real-time) and – based on prior related work on learning analytics (e.g., [5]) – a text-based console displaying further information about learner-editor interactions, e.g., manual use of the syntax validation.

Beyond these modes of data collection on learners' learning processes, we have implemented a software component to create and to conduct pre- and post-modeling online surveys of learners about their learning processes and demographic data within the modeling tool. Since learner-editor interactions are a rather restricted mode of observation of learning processes, we opted for additionally collecting verbal data by asking learners to think out load while modeling (e.g., [15]). Currently, a further component of the learning observatory is implemented to collect verbal data protocols by recording voice while learners are working on a modeling task. Generally, we view it as an open question whether and how to further enrich researchers' observations on learning processes beyond learner-tool interactions. As learning of conceptual modeling entails an intricate array of cognitive processes and performed actions, studying learners' learning processes suggests applying further complementary data collection approaches including, for instance, eye tracking (e.g., [16]) respectively video recording via webcam (e.g., [17]).

To evaluate the design and implementation, we are currently preparing for a small-scale study aimed at achieving first insights into barriers and difficulties faced in learning conceptual modeling and into learners' learning processes. Involving selected students working on a modeling task in the graphical user interface of the modeling tool, the design of the study utilizes the multi-modal data collection approach combining tracking learner-tool interactions, recording voice from learners and conducting pre- and post-modeling online surveys of learners. Informed by these preliminary insights, a large-scale study will be designed aimed at identifying patterns of learning processes, common learning barriers and learning difficulties.

References

1. Sedrakyan, G., Snoeck, M.: Cognitive Feedback and Behavioral Feedforward Automation Perspectives for Modeling and Validation in a Learning Context. In: Hammoudi, S., Pires, L., Selic, B., and Desfray, P. (eds.) Model-Driven Engineering and Software Development. MODELSWARD 2016. Communications in Computer and Information Science. pp. 70–92. Springer, Cham (2017)
2. Ortner, E.: Sprachbasierte Informatik: Wie man mit Wörtern die Cyber-Welt bewegt. Edition am Gutenbergplatz, Leipzig (2005)
3. Frank, U.: Outline of a Method for Designing Modelling Languages. ICB-Research Report No. 42, Institute for Computer Science and Business Information Systems (ICB), University Dusiburg-Essen, Germany (2010)

4. Soler, J., Boada, I., Prados, F., Poch, J., Fabregat, R.: A Formative Assessment Tool for Conceptual Database Design Using UML Class Diagram. International Journal of Emerging Technologies in Learning. 5, 3, 27–33 (2010)
5. Sedrakyan, G., Snoeck, M.: Effects of Simulation on Novices' Understanding of the Concept of Inheritance in Conceptual Modeling. In: Jeusfeld, M.A. and Karlapalem, K. (eds.) Advances in Conceptual Modeling. ER 2015. LNCS, vol 9382, pp. 327–336. Springer, Cham (2015)
6. Cosentino, V., Gérard, S., Cabot, J.: A model-based approach to gamify the learning of modeling. In: Jureta, I., Snoeck, M., Guizzardi, R., and Franch, X. (eds.) Proceedings of the 5th Symposium on Conceptual Modeling Education and the 2nd International iStar Teaching Workshop, SCME-iStarT 2017, pp. 15–24. CEUR-WS, Valencia, Spain (2017)
7. Boughzala, I., Chourabi, O., Lang, D., Feki, M.: Feedback on the integration of a serious game in the data modeling learning. In: Proceedings of the 50th Hawaii International Conference on System Sciences. HICSS 2017, pp. 735–742. Waikoloa Village, Hawaii, USA (2017)
8. Tsarmpou, P., Tambouris, E.: Using learning analytics to enhance UML use case diagrams assimilation in a distance education course. International Journal of Learning Technology. 10, 4, 274–290 (2015)
9. Sedrakyan, G., Snoeck, M., De Weerdt, J.: Process mining analysis of conceptual modeling behavior of novices - Empirical study using JMermaid modeling and experimental logging environment. Computers in Human Behavior. 41, 486–503 (2014)
10. Ericsson K.A., Simon H.A.: Verbal reports as data. Psychological Review 87, 3, 215–251 (1980)
11. Frank, U.: MEMO Organisation Modelling Language (2): Focus on Business Processes. ICB-Research Report No. 49, Institute for Computer Science and Business Information Systems (ICB), University Dusiburg-Essen, Germany (2011)
12. Ternes, B.: Design and evaluation of a web-based modeling platform to support the learning of conceptual modeling and of studying the corresponding learning processes. In: Gulden, J., Nurcan, S., Reinhartz-Berger, I., Guédria, W., Bera, P., Guerreiro, S., Fellmann, M., and Weidlich, M. (eds.) Proceedings of 8th International Workshop on Enterprise Modeling and Information Systems Architectures (EMISA) co-located with the 29th International Conference on Advanced Information Systems Engineering 2017 (CAiSE 2017), pp. 138–142. CEUR-WS, Essen, Germany (2017)
13. Ternes, B., Strecker, S.: A web-based modeling tool for studying the learning of conceptual modeling. In: Schaefer, I., Karagiannis, D., Vogelsang, A., Méndez, D., and Seidl, C. (eds.) Modellierung 2018, Braunschweig, Germany, pp. 325–328. Gesellschaft für Informatik e.V. (2018)
14. Knöpfel, A., Gröne, B., Tabeling, P.: Fundamental Modeling Concepts. John Wiley & Sons, Bell & Bain, Glasgow (2005)
15. Haisjackl, C., Barba, I., Zugal, S., Soffer, P., Hadar, I., Reichert, M., Pinggera, J., Weber, B.: Understanding Declare models: strategies, pitfalls, empirical results. Software and Systems Modeling. 15, 2, 325–352 (2016)
16. Zugal, S., Pinggera, J., Neurauter, M., Maran, T., Weber, B.: Cheetah Experimental Platform Web 1.0: Cleaning Pupillary Data, https://arxiv.org/abs/1703.09468 (2017)
17. Zitek, A., Poppe, M., Stelzhammer, M., Muhar, S., Bredeweg, B.: Learning by Conceptual Modeling – Changes in Knowledge Structure and Content. IEEE Transactions on Learning Technologies. 6, 3, 217–227 (2013)

Exergames & Dementia: An interactive System to for People with Dementia and their Care-Network

David Unbehaun[1] & Konstantin Aal[1]

[1] University of Siegen, Institute of Socio-Informatics, Siegen, Germany
{firstname.surname}@uni-siegen.de

Abstract. Dementia and its individual and social consequences can be positively affected by a range of activities that can postpone individual decline and the development of dementia. ICT can play a major role to have positive effects on people with dementia (PwD), but also on other stakeholders such as professional and informal caregivers. The aim of the presented system is to increase the physical and cognitive capabilities to cope with the activities of daily life (ADL) of PwD, as well as to relieve the caregivers and the social care-network. To this end, the MobiAssist System is being developed with and for people with dementia and their caregivers. The aim is to provide exercises (physically, cognitive and creative) to enhance mental and physical resources for ADL. Feedback from the participants were positive towards the general movement and cognitive capabilities, social and cross-generative interaction, and therefore a relief for informal and professional caregivers.

Keywords: Dementia; Exergame; ICT, Participatory Design; QoL; Care

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

Workshops:

Track Chairs:

Victoria Wenzelmann
Universität Siegen

Maximilian Krüger
Universität Siegen

Michael Ahmadi
Universität Siegen

Workshop Ethics and Morality in Business Informatics (Workshop Ethik und Moral in der Wirtschaftsinformatik – EMoWI'19)

Jens Gulden[1], Alexander Bock[1], Sergio España[2]

[1] University of Duisburg-Essen, Research Group Enterprise Modeling and Information Systems, Essen, Germany
{jens.gulden,alexander.bock}@uni-due.de
[2] Utrecht University, Department of Information and Computing Sciences, Utrecht, The Netherlands
s.espana@uu.nl

Abstract. The aim of the first edition of the EMoWI workshop was to establish a new forum for Business Informatics researchers and practitioners to reflect on the various ways in which the concern with business information systems and digital technologies gives rise to questions and issues with an ethical dimension. The several contributions of the workshop have made it plain that ethical questions indeed crop up in many fields of Business Informatics, ranging from specific research objects such as digital platforms to methodological pre-suppositions of the discipline at large. This chapter provides an overview of the background and the contributions of the EMoWI workshop 2019.

Keywords: Ethics, Values, Morals, Moral Philosophy, Business Informatics, Information Systems.

1 Introduction

According to Immanuel Kant, ethics is concerned with the question, "What ought I to do?", as opposed to the other two basic questions "What can I know?" and "What may I hope?" [1, p. 635, AK 833]. "What ought I to do as a Business Informatics researcher?" was therefore the leading question of the workshop *Ethics and Morality in Business Informatics* (EMoWI'19), held in its first edition at the 14th International Conference on Wirtschaftsinformatik (WI 2019), in Siegen.

The idea for the workshop emerged from the recognition that while the examination of ethical questions has received a fair amount of attention in specific branches of information systems research, computer science, and the engineering sciences (see, e.g., [2, Part V] [3]), these examinations have tended to be underrepresented in major Business Informatics conferences in recent years. Yet the need for ethical reflections is more pressing than ever, given that information systems and digital technologies pervade both private and professional life to an unprecedented degree. The many developments in digital technologies bring forward classical problems of moral

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

philosophy in new guises, as well as genuinely new species of situations in which good and bad, right and wrong conduct needs to be distinguished.

These developments pose significant challenges for the discipline of Business Informatics. Given that Business Informatics is conventionally expected to develop new ways of dealing with the emergent opportunities and challenges in a socially and technologically evolving society, the discipline cannot focus on the design of socio-technical information systems in an instrumental sense alone. Instead, Business Informatics also has the responsibility to develop new ways of meeting the diverse and complex ethical ramifications of the use and design of information systems in society with prudence and responsibility. Our motivation in organizing this workshop was to advance the study of such moral responses to a rapidly changing world, and to promote reflections on the consequences for Business Informatics in research, teaching, and the public sphere.

2 Central Themes and Challenges of Ethics in Business Informatics

The discipline of Business Informatics is rife with ethical questions and issues. In common with its principal subject—the understanding and development of socio-technical information systems—ethics and morality in Business Informatics can be studied from a variety of perspectives. These perspectives may concern the actual research objects of the discipline, or they may concern meta-scientific, methodological issues. On the level of concrete research objects, many areas of Business Informatics are bound up with ethical issues. For example, the design of artifacts and methods of Business Informatics has an ethical dimension, as the developed artifacts will typically have normative consequences for the behavior of people developing and operating business information systems. Likewise, the study of observable behavior between humans and software systems can investigate a diverse range of moral aspects, as information systems in many ways affect the actions, perceptions, and language of people involved with them. It is also worthy of examination whether conventional self-conceptions of professionals in Business Informatics embody moral, or perhaps morally problematic, premises. On a methodological level, in turn, ethical questions can concern basic assumptions and paradigms inherent in the research methods of Business Informatics.

It would be hopeless to attempt to provide a more exhaustive review of ethical questions in Business Informatics here. To make our case more tangible, we instead wish to highlight two general areas in which both classical and novel ethical questions arise in particular.

Decision Support in Information Systems. It is conventionally assumed that decision support is one of the basic tasks of computerized information systems. The history of Business Informatics is inextricably entrenched with Management Studies and Economics, where the leading conception of rational decision making has long been the expected utility model in different variants [4]. When applied to decision problems with an ethical dimension, this model naturally lends itself to a *consequentialist* conception of moral philosophy, famously advanced, for example, by Bentham [5] where alternatives are evaluated by the aggregated amount of pleasure or happiness obtained. However, such moral philosophies tend to neglect absolute rights and wrongs, such as the unalterable obligation not to kill. Such absolute principles are emphasized by *deontologist* moral theories, such as Kant's moral philosophy [6]. It is a question worth considering whether ordinary forms of decision support implemented in contemporary information systems implicitly adopt specific species of moral philosophies—and whether this has an effect on the moral evaluation of the courses of action recommended by these systems.

The Ethics of Developing and Using Business Informatics Artefacts. Humans and organizations shape the technology they design and implement, infusing their values into the resulting artefacts. In turn, those technological artefacts shape human experience and society. Value sensitive design is a theoretically grounded approach to the design of technology that systematically accounts for human values throughout the design process [7]. Neglecting the role of ethical and human values in business informatics product developments does not make those projects and products value-free. Moreover, we are witnessing a recent increase of interest in researching the role and effects of ethics and values in software engineering [8] [9]. Researchers formulate questions related to the prominent values of software engineers themselves [10], the ethical concerns of software development organizations (e.g. diversity, workplace quality), and how software development processes should account for ethical values. Business Informatics, with its focus on supporting business management and operations, offers additional research questions, such as how software can contribute to increasing enterprise responsibility [11].

Overall, it is evident that ethical questions are deeply rooted in basic presuppositions and research objects of the field of Business Informatics.

3 Contributions of the Workshop

The contributions to this workshop address a rich spectrum of the ethical and value-related problem areas in Business Informatics outlined above. Out of seven submissions, after careful review by our program committee members, we have selected five papers for presentation at the workshop and publication in the workshop proceedings.

Paper 1. In studying how information systems can be used to improve human welfare, it is instructive to consider how classical and contemporary philosophical traditions have answered related questions. The paper *"Doing Good Better: What We Can Learn from Effective Altruism"* by Alexander Herwix demonstrates how the core ideas of a recent movement with philosophical and anthropological roots, effective altruism, can be applied in the field of Business Informatics. Effective altruism, making the case for a deliberate and measured use of resources so as to generate the greatest human benefits possible, provides a number of practices and principles to help promoting good in different areas. Applying two core ideas from effective altruism, 'Shared Goals, Principles and Measures' and 'Cause Neutrality and Focus Area Selection', to the field of Business Informatics, the paper by Alexander Herwix proposes a number of reorientations for the discipline so that it may achieve more altruistic impact than it has had to date.

Paper 2. A central aim of Business Informatics is the design of new artifacts to support the development and use of information systems, commonly referred to as Design Science Research (DSR). Whether recognized or not, the design of artifacts is necessarily laden with ethical decisions. To promote awareness and evaluation of value-related design decisions in DSR projects the approach Value Sensitive Design (VSD) has been advanced in the literature. The paper *"Value Sensitive Design in Design Science Research Projects: The Cases of Affective Technology and Healthcare Technology"* by Oliver Heger reports on how VSD has been applied in two recent DSR projects and reflects on the lessons learned. The paper finds that VSD can indeed help identify, systematize, and assess value-laden design decisions and related conflicts in DSR projects. At the same time, the author stresses that more work is needed to obtain a better understanding of how approaches like VSD can raise Business Informatics researchers' awareness of ethical aspects in the design of socio-technical artifacts.

Paper 3. A view of the societal repercussions of the collaborative use of information technology and resulting ethical questions is developed by the paper *"Analyse der ethischen Fragestellungen bei der Konstruktion von digitalen Plattformen"* ("Analysis of Ethical Issues in the Construction of Digital Platforms") by Olga Levina. Considering the four dimensions Privacy, Access, Property and Accuracy, the paper reflects on ethical aspects of the governance structures of digital platforms. The examination demonstrates that over time an information asymmetry to the benefit of platform operators grows, which provides platform operators with instruments of power that potentially promote abuse. As a result, the paper argues for a socio-political debate about ethical questions around the construction of digital platforms.

Paper 4. Nudging has become a widespread mechanism in interactive and persuasive technologies, given its potential to influence the decision-making process of users and, eventually, modify their behavior. In the context of assistive technologies, nudging can be used to promote healthy behaviors and increase the independence of users (e. g., elders or people with an impairment). In their paper *"Moral Challenges in Modeling and Simulation of Behavior Change in Care"*, Stephanie C. Rodermund, Fabian Lorig, and Ingo J. Timm discuss the ethical implications of modeling and simulating nudging mechanisms and their corresponding user reactions. As they point out, such research and engineering practices entail ethical risks related to safety and privacy, among other concerns. Nonetheless, from a utilitarian perspective, once the risks have been analyzed and managed, the potential benefits in the well-being and independence of users makes nudging a convenient feature in assistive technologies.

Paper 5. The short paper *"Das Wertequadrat als Werkzeug der Wirtschaftsinformatik"* ("The Square of Values as a Tool in Business Informatics") by Alexander Rachmann addresses ethical questions on the methodical level of Business Informatics. It suggests to incorporate a conceptual tool called Square of Values into the methodical range of the discipline. The tool provides a systematic approach to find a balance between a pair of conflicting ethical values which both appear desirable. Together with an example of an ethical conflict in the domain of health-care oriented Smart Home applications, which targets the conflicting values of controlling a person's health status versus respecting a person's privacy, the paper suggests to connect the Square of Values with established modeling techniques for goals systems in Business Informatics.

As a whole, the contributions demonstrate that there is indeed much potential and need for research on ethical aspects in the objects and the methods of Business Informatics. Considering the range of topics discussed in the papers, it is clear that ethical questions are not placed at the margins of Business Informatics, but are instead entrenched in the very core subjects of the discipline.

4 Conclusions

The workshop *Ethics and Morality in Business Informatics* (EMoWI'19) was held in its first edition at the 14th International Conference on Wirtschaftsinformatik (WI 2019) in Siegen. We hope to have reinforced that the study of ethics and morality has an important place in Business Informatics, and that many areas in the field are worthy, and actually in need, of a careful study as to their moral implications. If anything, it can be expected that the ongoing digital transformation will further advance the impact, and thus the ethical relevance, that information systems will have for all people's life.

Many colleagues have contributed to the success of this first edition of the Workshop on Ethics and Morality in Business Informatics, so we would like to express our appreciation. We are grateful to the international program committee members, who have carefully reviewed the submitted papers. We sincerely thank the authors who were inspired by the topic of the workshop and have submitted papers for consideration; also, we thank those authors whose paper was accepted, for presenting their work in the workshop and participating in the lively debates. Last but not least, we would like to thank the organizing committee members of the WI 2019 conference for their support.

The workshop proceedings have been published at http://ceur-ws.org/Vol-2297.

References

1. Kant, I.: Critique of Pure Reason. Translated by Norman Kemp Smith. Macmillan, London (1929)
2. Meijers, A. (ed.). Philosophy of Technology and Engineering Sciences. Elsevier, Amsterdam, London (2009)
3. Bendel, O.: 300 Keywords Informationsethik. Grundwissen aus Computer-, Netz- und Neue-Medien-Ethik sowie Maschinenethik. Springer Fachmedien, Wiesbaden (2016)
4. Schoemaker, P. J. H.: The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. Journal of Economic Literature 20(2), 529–563 (1982)
5. Bentham, J.: An Introduction to the Principles of Morals and Legislation. In Two Volumes. A New Edition, Corrected by the Author. Vol. I. Pickering, London (1823)
6. Kant, I.: Groundwork of the Metaphysics of Moral. In: Gregor, M. Timmermann, J. (eds.) Immanuel Kant: Groundwork of the Metaphysics of Morals. Cambridge University Press, Cambridge, 1–72 (2012)

7. Friedman, B., Hendry, D. G., Borning, A.: A Survey of Value Sensitive Design Methods. Foundations and Trends in Human-Computer Interaction 11(2), 63–125 (2017)
8. Ferrario, M. A., Simm, W., Forshaw, S., Gradinar, A., Smith, M. T., Smith, I.: Values-first SE: research principles in practice. In: Companion Proceedings of the 38th International Conference on Software Engineering (ICSE 2016), pp. 553-562. ACM, New York (2016)
9. Aydemir F. B., Dalpiaz, F.: A Roadmap for Ethics-Aware Software Engineering. In: Proceedings of FairWare'18: IEEE/ACM International Workshop on Software Fairness, Gothenburg, Sweden, May 29, 2018 (FairWare '18). ACM, New York (2018)
10. Winter, E., Forshaw, S., Ferrario, M. A.: Measuring human values in software engineering. In: Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18), Article 48. ACM, New York (2018)
11. España, S., Brinkkemper, S.: Responsible software: a research agenda to help enterprises become more sustainable. In: 4th International Conference on ICT for Sustainability (ICT4S 2016), Amsterdam, pp. 141-150. Atlantis Press, Paris (2016)

Model-Based Compliance in Information Systems – Foundations, Case Description and Data Set of the MobIS-Challenge for Students and Doctoral Candidates

Constantin Houy, Jana-Rebecca Rehse, Martin Scheid, Peter Fettke

Institute for Information Systems (IWi) at the German Research Center for Artificial Intelligence (DFKI) and Saarland University
Campus, Building D 3_2
66123 Saarbrücken, Germany
firstname.lastname@dfki.de

Abstract. Information systems (IS) can significantly support the organization of business processes. However, the proceeding digitalization of processes can also lead to an increasing organizational complexity and the need to more intensely investigate the adherence to external or internal compliance rules. Process-related data from IS and underlying process models can, however, also contribute to an effective compliance checking. This paper first presents conceptual foundations of model-based compliance checking that motivated the MobIS-Challenge workshop for students and doctoral candidates at WI 2019. Second, we introduce the challenge itself and its corresponding data set. The data describes an exemplary travel management process in a medium-sized consulting company and served for the development and validation of adequate solutions addressing the compliance checking requirements. Solutions accepted for presentation at the workshop are briefly outlined in this paper.

Keywords: GRC, Governance, Risk and Compliance, BPM, Business Process Management, Process Mining, Data Set

1 Introduction

Nowadays, business processes are increasingly carried out digitally with the help of information systems (IS), which can significantly support an effective, efficient and flexible management of business processes. However, the proceeding digitalization of business processes can also lead to an increasing complexity of organizations and the need to more intensely analyze and ensure the adherence to externally or internally specified compliance rules. In this context, process data from IS and underlying business processes as well as business process models can, however, also considerably contribute to an effective compliance checking. The usage of data from IS allows for an easier reconstruction of business processes, e.g. based on built-in logging mechanisms, and furthermore facilitates the identification of violations of internal or external compliance rules. In this context, there are some research streams,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

such as process discovery and conformance checking, that develop new methods and techniques to analyze the process data logged by information systems and to use the gained insights for the benefit of the company.

This contribution presents the WI 2019-Workshop "MobIS-Challenge for Students and Doctoral Candidates: *Model-Based Compliance in Information Systems*", the topic's conceptual foundations, the use case which was investigated by the challenge participants as well as the corresponding data set. Participants of the MobIS-Challenge were supposed to identify opportunities to use IT tools (e.g. existing process mining tools, BPM solutions, self-developed programs etc.) for analyzing and improving the business process compliance in the addressed business travel management scenario, specifically by examining the compliance of this process and pointing out its weaknesses.

The remainder of this paper is organized as follows: in section 2, we introduce some conceptual foundations of model-based compliance management with a particular focus on business process data from information systems. In section 3, we describe the travel management use case to be analyzed in more detail providing a verbal description as well as an according business process model which served as a basis for the simulation and development of the data set. Section 4 provides a detailed description of the data set focusing on the most severe compliance violations included in the approximately 6,500 business travel management cases contained in the data. Section 5 presents the tasks which had to be addressed in the challenge as well as a brief description of potential solutions by the authors of this paper, while section 6 introduces accepted solutions which participants submitted to the MobIS-Challenge.[1] Section 7 concludes the paper.

2 Business Process Compliance

Business process models can serve as an instrument to express and clarify the course of activities in the context of value creation in organizations [1]. While business processes can be understood as sequences of executions for the purpose of creating goods and services [2], business process models are representations of business processes which provide the basis for several different tasks of Business Process Management (BPM) [3], such as process implementation, execution, controlling or systematic process improvement [4].

In order to support their daily operations, business organizations use information systems (IS), like enterprise systems (ES) for enterprise resource planning (ERP), supply chain management (SCM) or customer relationship management (CRM) etc. Such IS – no matter whether they are process-oriented and explicitly produce so-called process log data or not – generate data, which can serve to obtain a view of the underlying business processes. The data generated by IS can, thus, also serve for the identification of compliance violations.

[1] We currently plan to separately publish a detailed report on the accepted solutions submitted to the MobIS-Challenge.

In literature, *compliance* is one major aspect of the comprehensive topic addressed by the umbrella term *Governance, Risk and Compliance* (GRC). GRC and its related policies and rules as well as technical support approaches and methods are supposed to ensure a good, responsible and sustainable management of organizations, which follow the applicable law and commonly accepted standards [5, 6]. Compliance management is supposed to ensure the conformity of "business processes, operations and practice […] with a prescribed and/or agreed set of norms" [7]. In this context, external and internal compliance requirements can be differentiated. Typical external compliance requirements are legal initiatives like the *Sarbanes-Oxley Act* (SarbOx) in the US or *Basel III* in the financial sector as well as the so-called *Bribery Act 2010* as an anti-corruption legislation example passed in the UK. Furthermore, many organizations have defined internal compliance requirements, which have not been formulated by external authorities, but which are supposed to ensure a voluntary conformity of the organization's behavior with common standards.

The term *business process compliance* is ambiguous and used to address different concepts in literature. Some contributions aim at checking the compliance of business process instances (*as-is processes*) in terms of a defined process model, focusing on their identity without looking at operational business issues. Other contributions refer to business process compliance as a means of checking the operational compliance of an organization based on the underlying business processes, e.g. using process logs to identify and investigate potential violations of external or internal compliance rules [8-12]. In this contribution, we use the term *business process compliance* according to the latter understanding.

FELLMANN and ZASADA give a comprehensive overview of the current state-of-the-art in their review contribution investigating a total of 84 business process compliance approaches [13]. They identified different dimensions for compliance checking (p. 5) which are described in Table 1.

Table 1: Dimensions of Compliance Checking (Fellmann and Zasada, 2014)

Nr.	*Dimension*	*Sub-Dimensions*
1	*Scope*	Order and occurrence, Information, Resource, Time, Location
2	*Lifecycle phase*	Design, Execution, After Execution
3	*Formality*	Verification / Validation, Business-oriented
4	*Contribution type*	Technical artefact, Method, Other

Known approaches for a business process-oriented compliance checking, e.g. use process mining techniques [14] or so-called control patterns [15]. Furthermore, there are several commercial tools, which already implement compliance checking techniques in different contexts [16]. The following section describes the underlying business travel management case treated in the MobIS-Challenge.

3 Case: Business Travel Management Process

3.1 Case Description and Process Model

The case that we provide for the workshop describes a business travel management process in a medium-sized software consulting company. While the data itself was generated by simulation, the process and its governing compliance rules are inspired by one we have encountered in a recent research project.

The goal of the business travel management process is to keep track of all business trips the employees take and their related expenses, such that they can be invoiced to the respective customer, for whose project the respective trip is taken. As is usual in consulting companies, the employees travel quite often to meet with customers, but as software consulting includes some work that can be done remotely, they are not constantly traveling. In order to better control the bookings, to take advantage of economies of scale in the booking process, and to avoid lengthy reimbursements, the company has decided to install a separate travel department, where multiple travel agents are responsible for booking business trips, always in accordance with the respective employee.

To improve the internal process organization, the company has developed its own internal workflow management system which can be accessed by each employee. Travel management is fully covered and logged by this system, with the travel management process implemented as a workflow and the tasks and rights assigned according to the employee's role in the company.

Within the process, there are four acting roles:

(1) the Employee, who wants to go on a business trip,
(2) the Manager, who has to approve the trip and the expense report,
(3) the Travel Department, which is responsible for bookings and price information, and
(4) the Accounting, which is responsible for calculating and reimbursing costs.

In our case company, there are 300 employees, 15 managers (including 3 directors), 5 travel agents, and 10 accounting clerks.

The process starts, when an employee files a travel request. Such a request offers two options: The employee can either directly file a request or she can initiate a preliminary price inquiry, which is helpful if the travel expenses and potential booking options are unclear. In this case, the request is forwarded to the travel department, where a travel agent provides a booking proposal and discusses it with the employee. If the employee accepts this proposal, she can adapt her price inquiry accordingly and then transform it into an official travel request. If the proposal is not accepted, she first has to check whether the trip is still necessary and all data is up-to-date, before requesting an update of the booking proposal from the travel agent.

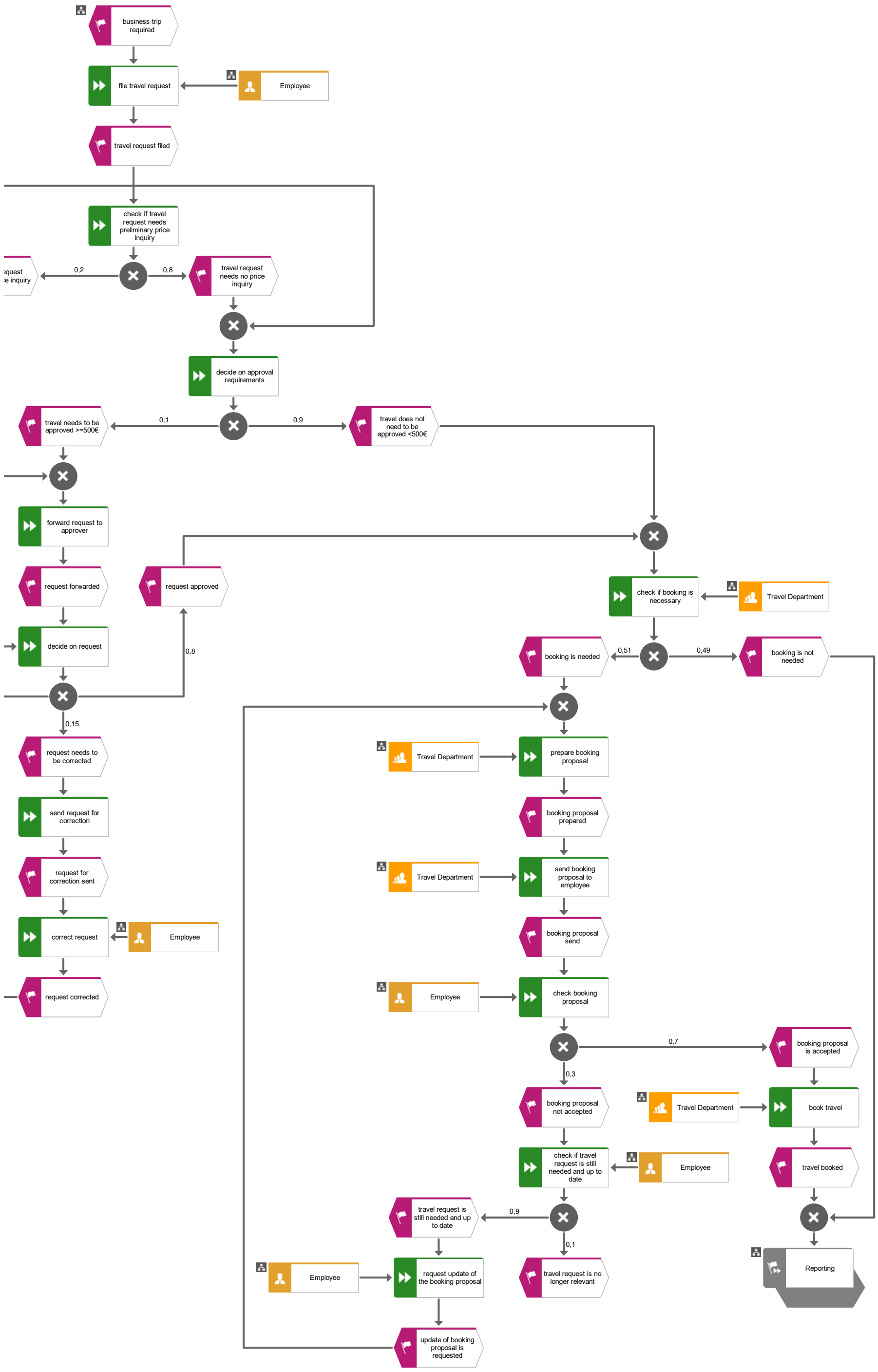


Figure 1: Excerpt of the *planning part* of the travel management process

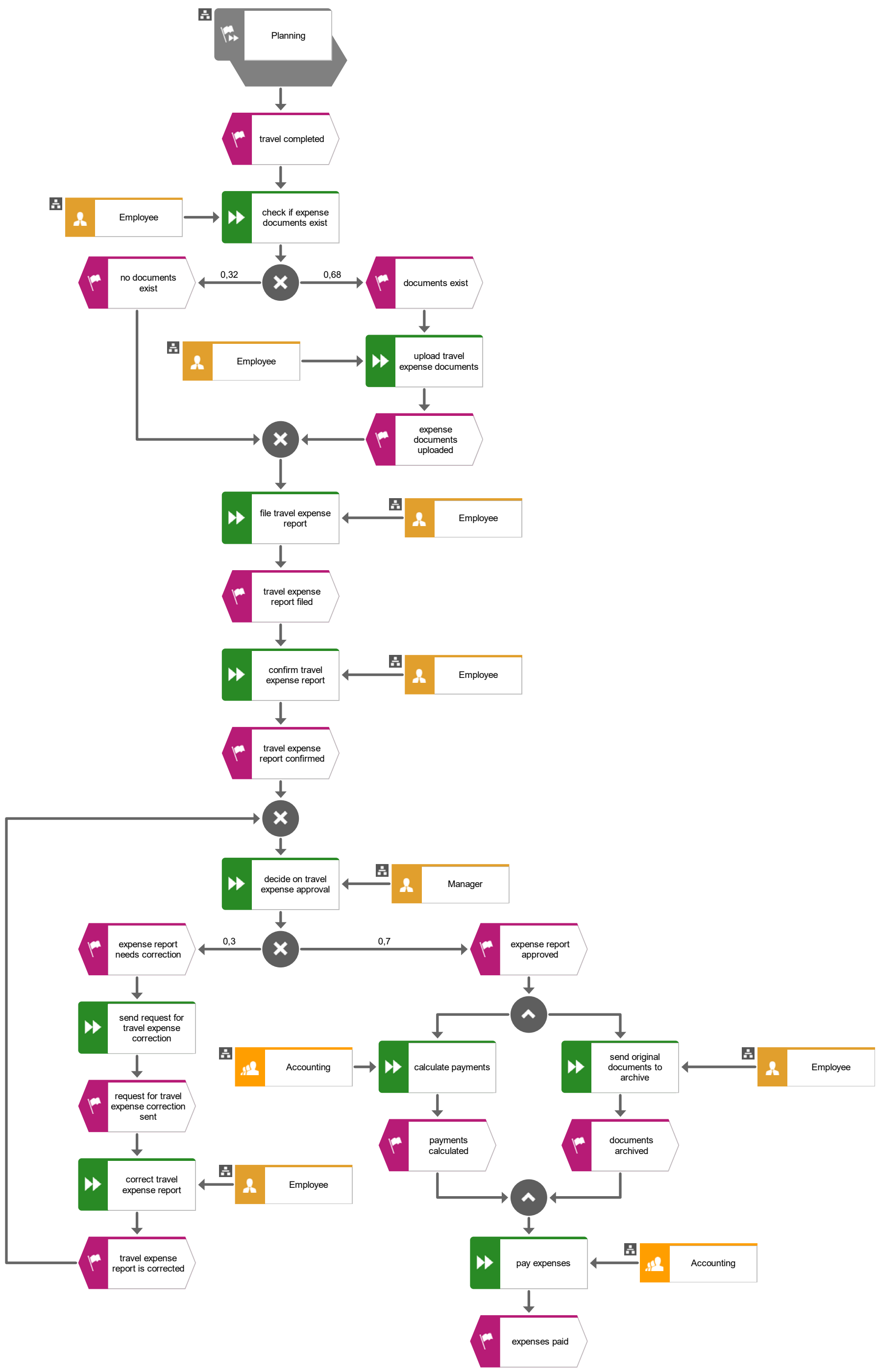


Figure 2: The *reporting part* of the travel management process

When the travel request is officially filed, it has to be checked for approval before it can be handled by the travel department. The workflow system automatically checks whether the request fulfills the approval requirements and forwards it to the responsible manager for approval. The manager checks the request and either approves it, rejects it, or asks for a correction. In the latter case, the system redirects the request to the employee, such that she can make adjustments according to the manager's requests. This correction process is repeated, until the manager finally approves or rejects the travel request.

Once a travel request is approved, or if approval is not required, it is forwarded to the travel department and assigned to a travel agent, who checks whether the request requires any bookings. If not, for example if the employee takes her own car or a company car for a business trip, there is nothing left to do for the travel department. If yes, the travel agent prepares a booking proposal according to the employee's specifications and sends it back for approval. If the employee approves the proposal, the travel agent confirms and pays for all bookings (e.g. hotel, flights, or rental cars). If the employee does not approve the proposal, she has to confirm the data and relevance of the travel request, before she can ask the travel department for an updated booking proposal.

After a business trip is concluded, the employee has to fill out a travel expense report in order to be reimbursed for any travel-related costs. To ensure correct accounting procedures, employees also have to fill out a report if no expenses have incurred. Therefore, the employee first needs to check whether she has any travel-related expense documents (e.g. invoices or receipts). If such documents exist, they have to be uploaded in a digital form, such as a scan. Afterwards, the employee fills out the travel expense report as provided by the workflow system. The confirmed report is automatically forwarded to the respective manager for approval. If the manager decides that the report cannot be approved, it is sent back to the employee for corrections. After the report is approved, the accounting department is in charge of calculating the total travel costs, archiving the travel-related documents, and paying the expenses of the employee.

3.2 Compliance Rules

Within the travel management process, there are a number of external and internal compliance rules, which must be followed. The internal rules are mainly important to keep the accounts correct and up-to-date, whereas the external rules are necessary for invoicing travel expenses to the customer. In detail, the travel management process conforms to the following compliance rules:

(1) For each business trip, an according travel request must be filed and, if necessary, approved before the beginning of the trip.
(2) Business trips must be necessary to ensure the success of a project. If this necessity cannot be documented, a manager might reject the request.
(3) A travel request must contain realistic cost estimations. The real costs in the travel expense report filed after the trip should not exceed this estimation.

(4) If the estimated travel cost does not exceed 500€, it does not need to be approved.
(5) If the estimated travel cost exceeds 500€, the trip must be approved by the employee's responsible manager.
(6) Managers' trips must be approved by a director. The three directors approve each other's trips.
(7) Managers should promptly approve, reject, or react to incoming requests.
(8) If possible, trips should be booked and paid for by the travel department.
(9) After a trip has ended, the travel expense report should be filed immediately.
(10) All travel-related expenses should be documented with a receipt.
(11) Travel expense reports must be approved by the employee's responsible manager. Again, managers' travel expense reports are approved by a director and the directors approve each other's reports.
(12) Managers should check requests and expense reports carefully and ask for corrections, if they find any rule violations.

4 Data Set[2]

4.1 Model Development

In order to generate a viable process log for the MobIS-Challenge case, we first developed a process model that we could use as the basis for data generation. We used the ARIS simulation component for data generation and modelled our process as an *Event-Driven Process Chain* (EPC). Then we enriched it with data that is required for simulation. According to the description above, we separated the process into two major parts. Large parts of the travel planning process are shown in Figure 1. While the price inquiry handling is described on the left side of the complete process (not displayed in the excerpt in figure 1), the right side of the process model deals with approving travel requests in the loop on the left and handling the bookings in the subprocess on the right. The second part of the process, expense reporting, is shown in Figure 2. First, the employee uploads all travel-related documents and produces an expense report, which is then approved by the manager in the loop on the bottom left. Expenses are reimbursed by the accounting department on the bottom right.

ARIS offers a multitude of attributes to be defined for each model element, some of which were necessary to ensure that our simulation would produce a viable data set. In our case, we needed to define an executing role for each function to assign resources in the process log, the number of employees that instantiate each role, probabilities for each XOR-connector to determine the path frequency, execution times for each function to allocate sufficient time for its execution, and schedules for employees to account for normal working hours. We also associated the start event

[2] Data set source: Scheid, M., Rehse, J.-R., Houy, C., & Fettke, P. (2018). *Data Set for MobIS Challenge 2019 [Data set]*. https://doi.org/10.13140/RG.2.2.11870.28487.

with an instantiation schedule and a fluctuating delay to introduce some randomness into the start times of each case.

As can be seen in the process model, each function is associated with a role that is responsible for its execution; roles without an explicitly associated role are automatically executed by the workflow system itself. Each role is associated with a schedule, which determines its typical working hours. Employees can work anytime between 6am and 11pm, which factors in that consultants sometimes do organizational tasks like travel management at odd hours. The administrative personnel, i.e. the travel agents and accounting clerks, work typical office hours from 9am until 5pm. We included one week of vacation time in July, where no travel agent was working. In the managers' schedule, we defined their working hours to be only between 1pm and 2pm to account for the fact that managers are typically very busy and only tend to administrative tasks like travel management at certain hours, e.g. after lunch.

We also defined a fluctuating execution time for each function. ARIS allows several options to set either strict or varying execution times and we decided to model them all as a capped normal distribution, specifying the expected value, standard deviation, minimum and maximum time for each function. Realistic values were set for each function, such that automated functions executed by the workflow system only take a few seconds, whereas time-intensive functions like preparing a booking proposal takes several hours.

The number of employees (300) and managers (15) in the company were predefined before simulation, but the necessary number of travel agents had to be determined empirically, such that some, but not too many cases were piling up at any given time. 5 travel agents turned out to be a sufficient number.

For most functions, we did not have to define static waiting times to account for employees being busy with other tasks that have nothing to do with travel management. All waiting times for the travel department are caused dynamically, because other cases are handled first. Only the accounting functions wait statically for a few days to account for other responsibilities of the accounting department.

4.2 Data Generation

After the model was developed, we used it as the basis for simulating the process data. This simulation consisted of multiple steps.

Process Simulation with ARIS. Based on the developed process model, we used the ARIS simulation component to generate execution data. To account for the travel time that occurred between the two process parts, we connected them with an artificial function ("travel"), with fluctuating static waiting times (to account for the time passing between a travel request and the trip itself) and execution times (to account for the duration of the trip). We wanted to simulate data for one year, so the simulation period was set from October 1st 2016 to December 31st 2017, with the first 91 days functioning as a warm-up phase to have plenty of cases in the system. ARIS used the specified process data to simulate its execution. Each simulation took about

15 to 20 minutes to complete. We exported the case data from ARIS and converted it into a CSV file to proceed further.

Generating Additional Data with Excel. The ARIS simulation was only able to generate the process steps itself, so we had to enrich the log with additional data on travel costs and the organizational structure. We defined the company's internal organizational structure, such that we could assign each case to an employee and the responsible manager. There were three cost values that had to be generated, the estimated travel costs, the real travel costs, and the reimbursed costs. The estimated costs were calculated randomly, depending on the length of the trip and whether or not the travel request has to be approved. The real travel cost was calculated to randomly fluctuate in both directions around the estimated cost. Finally, the reimbursed costs depend on whether the travel department booked the trip for the employee. If yes, they were lower than the travel costs, if no, the two numbers were equal.

Manual Data Cleaning. After all data was generated, we had to manually go over it to remove some mistakes and irregularities, such as business trips during Christmas time. After cleaning the data and introducing compliance violations (explained in the next section) our final data set contains 6,555 cases with 26 activities and a total of 83,256 events.

4.3 Violations of Compliance Rules

After the process log was simulated, enriched, and cleaned, we introduced compliance violations. From a data perspective, there are two types of compliance violations, those that were caused by simulation parameters and were already present in the log and those that had to be entered manually.

Violations caused by Simulation Parameters. We defined our simulation parameters (schedules, waiting times, cost calculations) such that we deliberately build some compliance violations directly into the log. They are listed in Table 2 and shortly explained in the following.

Table 2: Compliance violations caused by simulation parameters

ID	*Compliance violation*	*Simulation cause*
1	Long delays in manager's reaction	Managers are encouraged to answer promptly to incoming travel requests, to avoid not reacting in time for the trip. (*Rule 7*)
2	Long delays in expense reporting	Accounting should promptly reimburse the employees for their expenses, to avoid unnecessary payment legacy. (*Rule 9*)
3	Real travel expenses significantly exceed calculated expenses	Employees should give a realistic estimation of expected travel costs, to ensure necessary approvals and facilitate accounting. (*Rule 3*)

Violation (1) stems from a time restriction. According to their time plan, managers spend one hour each day for administrative tasks, causing requests to build up. This artificial restriction causes a bottleneck in the process, such that travel requests are delayed for several days before being approved or declined. Violation (2) can be accounted to fluctuating static waiting times, which we attributed to the functions to simulate other responsibilities of the accounting department. Finally, violation (3) is due to the fact that during cost simulation, we had real travel expenses fluctuate in relation to calculated expenses, such that they sometimes are much higher.

Manually entered Compliance Violations. However, most compliance violations (especially those that deviated from the normal process flow) could not be built directly into the log, but had to be manually entered. These violations, which include 9 of the 12 in total, are listed in Table 3. For each violation, we explain how it contradicts our compliance rules and give its frequency in the log, i.e. the number of existing cases which we altered to violate compliance in the described way.

Table 1: Manually entered compliance violations

ID	*Compliance violation*	*Compliance explanation*	*Freq.*
4	Travel request is submitted after the trip	Travel request must be filed and approved before the trip. (*Rule 1*)	8
5	Only price enquiry is submitted	Price enquiry must be converted into a travel request before the trip. (*Rule 1*)	12
6	Manager approves his own trip	Manager's trips must be approved by a director (four eyes principle). (*Rule 6*)	5
7	Manager approves his own expense calculation	Manager's expense calculations must be approved by a director (four eyes principle). (*Rule 6*)	10
8	Trip is approved by the wrong manager	Trips must be approved by the employee's own manager. (*Rule 5*)	11
9	Employee travels despite rejected travel request	Travel requests must be approved by a manager to ensure their necessity. (*Rule 1+2*)	2
10	Multiple travel requests (less than 500€) for one trip	Requests only have to be approved if the estimated costs exceed 500€. (*Rule 4+5*)	8
11	New travel request after rejection	The manager assessed the trip as unnecessary for the project success. (*Rule 2*)	3
12	Paid expenses exceed calculated expenses	The accounting department paid more to the employee than costs incurred for the trip. (*Rule 3*)	17

The following two sections give a brief description of the tasks and *potential* solutions to the problem to be treated in the MobIS-Challenge (section 5) and, furthermore, present the *submitted* solutions which were accepted for presentation at the MobIS-Challenge (section 6).

5 Tasks and Potential Approaches

The participants of the MobIS-Challenge were supposed to identify opportunities to use IT tools (e.g. existing process mining tools, BPM solutions, but also self-developed programs) for analyzing and improving process compliance and pointing out process weaknesses, even beyond conformance issues.

The tasks and leading questions for the MobIS-Challenge were as follows:

(1) *Describe the process depicted in the log with graphical means. From this description, derive meaningful compliance rules that go beyond the [.] description [given in the call for papers].*

(2) *Which compliance violations can be found in the data? How can these be prevented?*

(3) *Beyond compliance issues, which weaknesses in the process or the organization can be found in the data? How could these be improved?*

(4) *Which additional insights can be drawn from the data? You can use any tool to develop interesting additional insights in a creative way.*

To address and answer these questions, different approaches can be used. At first, the data should be carefully surveyed in order to get an overview of available data fields and their meanings in particular contexts. There is, e.g. a significant difference between the meanings of the different values in the field "cost" depending on the process step in which the costs play a role. In this context, we can differentiate between (a.) estimated costs, (b.) real costs and (c.) reimbursed costs.

Common *process discovery* approaches can be used to derive a graphical process model from the provided data set. In this context, different discovery algorithms (*heuristics miner*, *fuzzy miner*, *alpha miner* etc.) can be used and the different results can be compared. An adequate choice of the settings concerning activities and traces should be elaborated in order to develop a helpful process model which supports finding answers and solutions to the above problems.

Next, most process mining tools offer the opportunity to calculate distributions of cases, the identification of outliers and an overview of different process variants in the data set. Furthermore, time-related analyses can be conducted, e.g. to identify anomalous process durations.

Furthermore, there are several different possibilities to perform advanced analyses by identifying clusters in the data representing different classes of process variants as well as outliers which can then be analyzed in more detail to identify function deviations or anomalies. Moreover, the participants of the MobIS-Challenge could also "manually" develop a process model based on the data, e.g. a petri net, and then perform an automated conformance checking against the data set in order to identify outliers and anomalies. The same can be done in terms of rules. Obvious compliance rules which were followed in data set can be modelled and then be checked against the whole data set to identify outliers. The following section presents the solutions accepted for presentation at the workshop.

6 Solutions Accepted for Presentation at the Workshop

The following submitted solutions were accepted for presentation at the 2019 MobIS-Challenge workshop at WI 2019 in Siegen:

(1) In their extended abstract submission *Conformance Checking with Dynamic Condition Response (DCR) Graphs – An application to the MobIS Challenge 2019*, DUNZER, BAIER and STIERLE use a process conformance checking technique based on so-called *Dynamic Condition Response* (DCR) graphs [17]. Their approach follows the declarative process modeling paradigm and was instantiated as a software tool using Python. In order to answer the questions and fulfil the tasks of the MobIS-Challenge the CRISP-DM framework was used as a basis for structuring the project. The DCR graphs serve for the formulation of compliance rules which are then used for the identification of compliance violations in the data set.

(2) In their extended abstract submission named *Detection of Compliance Rule Violation in Business Processes using Sequence-to-Sequence Autoencoder*, WILLEMS and PFEIFFER describe the development of a framework for an automated detection of compliance violations based on neural networks. A sequence-to-sequence long short term memory (LSTM) autoencoder was trained using the entire event log from the MobIS-Challenge data set. By encoding and decoding traces from the event log, differences between input and output traces which are likely to indicate anomalous behavior can be calculated. The reconstruction error is used to find sequences of suspicious events. Furthermore, the extended abstract reports on the usage of state-of-the-art process mining tools like ProM, Disco, Celonis and bupaR for the further analysis of the event data set in order to provide detailed process mining-related analyses in the planned completed report.

7 Conclusions

This contribution presented the WI 2019-Workshop "MobIS-Challenge for Students and Doctoral Candidates: *Model-Based Compliance in Information Systems*". We provided a detailed description of the use case, which had to be investigated by the challenge participants as well as the corresponding data set. We have accepted two interesting initial descriptions of solution approaches. The challenge participants were invited to present detailed information concerning their solutions at the workshop.

Beyond the workshop in Siegen at WI 2019, we believe that the provided data set can also serve for and support further BPM research endeavors, e.g. in terms of the validation and evaluation of process mining or data analytics approaches for the investigation of data, which is relevant for business process compliance issues.

References

1. Curtis, B., Kellner, M.I., Over, J.: Process Modeling. Communications of the ACM 35, 75–90 (1992)
2. Scheer, A.-W.: ARIS – Business Process Frameworks. 3. edn. Springer, Berlin (1999)
3. van der Aalst, W. M. P.: Business Process Management: A Comprehensive Survey. ISRN Software Engineering, 1–37 (2013)
4. Houy, C., Fettke, P., Loos, P.: Empirical research in business process management – Analysis of an emerging field of research. Business Process Management Journal 16, 619–661 (2010)
5. Schäfer, T., Fettke, P., Loos, P.: Towards an Integration of GRC and BPM – Requirements Changes Caused by Externally Induced Complexity Drivers. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM 2011. Workshops Part II. LNBIP, vol. 100, pp. 344–355. Springer, Berlin (2011)
6. Becht, M., Bolton, P., Roell, A.: Corporate governance and control. In: Constantinides, G. M., Harris, M., Stulz, R. M. (eds.) Handbook of the Economics of Finance, pp. 1–109. Elsevier (2003)
7. Sadiq, S., Governatori, G.: A Methodological Framework for Aligning Business Processes and Regulatory Compliance. In: vom Brocke, J., Rosemann, M. (eds.) Handbook of Business Process Management 2. Strategic Alignment, Governance, People and Culture, pp. 159–176. Springer, Heidelberg (2010)
8. Schultz, M.: Business Process Compliance from an Audit Perspective. Universität Hamburg (2015)
9. Seeliger, A., Schmidt, B., Nolle, T., Mühlhäuser, M.: Process Compliance Checking using Taint Flow Analysis. In: International Conference on Information Systems (ICIS), Dublin, Ireland (2016)
10. Delfmann, P., Hübers, M.: Towards Supporting Business Process Compliance Checking with Compliance Pattern Catalogues – A Financial Industry Case Study. Enterprise Modelling and Information Systems Architectures 10, 67–88 (2015)
11. Schultz, M.: Enriching Process Models for Business Process Compliance Checking in ERP Environments. In: vom Brocke, J., Hekkala, R., Ram, S., Rossi, M. (eds.) Design Science at the Intersection of Physical and Virtual Design, LNCS, vol. 7939, pp. 120–135. Springer (2013)
12. Morana, S., Schacht, S., Scherp, A., Maedche, A.: Designing a Process Guidance System to Support User's Business Process Compliance. In: International Conference on Information Systems (ICIS), Auckland, New Zealand (2014)
13. Fellmann, M., Zasada, A.: State-of-the-Art of Business Process Compliance Approaches: A Survey. In: European Conference on Information Systems (ECIS), Tel Aviv, Israel (2014)
14. Caron, F., Vanthienen, J., Baesens, B.: Comprehensive rule-based compliance checking and risk management with process mining. Decision Support Systems 54, 1357–1369 (2013)
15. Schaefer, T., Fettke, P., Loos, P.: Control Patterns – Bridging the Gap between IS Controls and BPM. In: European Conference on Information Systems (ECIS), Utrecht, Netherlands (2013)
16. Kochanowski, M., Drawehn, J., Kötter, F., Renner, T.: Compliance in Geschäftsprozessen. Business Process Management Tools 2014. Fraunhofer, Stuttgart (2014)
17. Mukkamala, R. R.: A Formal Model For Declarative Worklflows – Dynamic Condition Response Graphs. PhD thesis. IT University of Copenhagen (2012)

Report of the Workshop on Concepts and Methods of Identifying Digital Potentials in Information Management

Stefan Koch[1], René Riedl[1,2], and Manuel Mühlburger[1]

[1] Johannes Kepler University, Linz, Austria
{stefan.koch,rene.riedl,manuel.muehlburger}@jku.at
[2] University of Applied Sciences Upper Austria, Steyr, Austria
{rene.riedl}@fh-steyr.at

Abstract. Recently a discussion has been instigated in the scientific community on whether our understanding of information management (IM) is still up to date [1]. A workshop on the concepts and methods of an additional IM perspective which intends to identify possible technology applications (by putting the primary focus on the possibilities offered by information technology rather than on the support of existing organisational functions and task structures) was held at the WI Conference 2019. This paper summarizes the conceptual input of the workshop and provides an overview of the initial workshop contributions.

Keywords: Digitalization, Digital Transformation, Information Management,

1 Workshop Topic

The changing understanding of information management (IM) towards a more enabling and proactive instead of a supportive and reactive organisational function is creating pressure for practitioners to deliver value from IT use in organisational contexts [2]. The rise of the now omnipresent terms of digital transformation (DT) and digitalization have contributed to this increasing pressure and have also caused increased expectations regarding the IT function's contribution to organisational success. Practitioners are expected to instigate, plan, and execute DT projects which do not primarily focus on cost reduction and productivity increases, but also contribute to new innovative business model transformation. As a consequence of this development, the scientific community is expected to advance existing and to develop new models, frameworks and methods which describe and explain technology driven changes in organisational value creation. These concepts should enable practitioners to rethink and refine existing organisational value creation structures by more directly considering the opportunities provided by digital technologies.

Following the aspect of increasing market dynamics, a fast organisational reaction capability to technological innovations has been identified as a major success factor for organisations [3]. This organisational agility requires organisations to not only be aware

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

of technological possibilities and advancements, but also to identify their potentials for the organisation [4]. Understanding how these potentials can be identified and exploited by organisations is therefore a necessity for a modern and successful IM.

Existing IM perspectives, methods and frameworks have a traditional focus on the organisational integration of information technology by following defined business strategies [5]. Although early works identified the potential of a proactive and enabling IM [6], these works remained largely unconsidered in practice. In essence, despite opposing calls in the academic literature [7, 8], IM in practice often concentrated on fulfilment of a supportive role. Whilst such as approach enables organisations to implement and operate new IT systems once their potential is identified, they give little guidance on how different application scenarios of new technology can be identified.

Whilst the tasks of this classical support role of IM will not lose relevance in the future (because reliable systems are the foundation of successful business processes [9]), there is a growing need for organisations to identify new technology potentials [10]. The question by which mechanisms, structures, processes and capabilities organisations can efficiently identify potentials has not been a primary focus of the IM field, but is central if IM should fulfil a proactive and enabling role [10]. An additional perspective on IM which integrates organisations into a digital world, instead of integrating digital technologies into organisations, provides value to both academia and practice.

To contribute to this recent development, a workshop was held at the WI conference 2019. Specifically, this workshop focussed on the identification of digital potentials as a primary aspect of future IM, both from a conceptual and methodological perspective. The initial contributions to this workshop put the focus on three main methodological and structural aspects of the identification of digital potentials.

2 Overview of Workshop Contributions

The first contribution by Florian Rittmeier presents a research proposal and design for the development of an assistance system enabling organisations to identify digitalization potentials within existing business process structures. The author proposes a system supporting organisations specifically in the phases of process discovery, analysis and redesign. By specifying key process characteristics relevant for a later potential analysis, the proposed system provides support within the phase of process discovery. The proposed system would further support process analysis by identifying deficiencies within the process model based on formalized process weakness patterns. Based on the identified weakness patterns, a formalized set of recommendations is presented to further support the phase of business process reengineering. By combining weakness patterns and relevant recommendations into opportunity patterns the author intends to enable organisations to identify and restructure their processes in a more efficient and effective way that is less dependent on the experience and interpretation of individual process analysts. We understand the attempt of creating a formalized pattern set of digitalization opportunities within business processes as a promising approach which provides a valuable contribution to this workshop regarding methodological aspects.

The second contribution authored by Barbara Krumay, David Rueckel and Sabrina Schwarzgruber discusses aspects of assessing and evaluating technology potentials in agile environments. As the authors put forward, traditional assessment methods tend to limit the dimensions used to assess the value of IT-investments to certain short-term monetary aspects. The authors argue that in agile environments with fast changing organisational goals and structures more flexible assessment methods including additional levels of analysis are required. Real options are identified by the authors as a promising methodology to focus assessments on additionally created technology potentials. Furthermore, the authors discuss the possible value of impact assessments as a method for identifying and assessing additional potentials resulting from certain IT-investments. Impact assessments are used in a variety of disciplines as a methodology for identifying possible consequences of certain events. Whilst currently predominantly applied in environmental, social and legal contexts, the authors argue for an adaptation and adoption of such instruments in a technology context. We understand the approach of utilizing impact assessments and real options for the identification, evaluation and assessment of technology potentials as a valuable contribution to this workshop to complement existing methods.

The final contribution from Friedrich Holotiuk discusses the role of digital innovation labs (DILs) in enabling organisational ambidexterity. Ambidexterity describes managing the trade-off between exploration and exploitation activities and is traditionally conceptualized in the two different types of structural and contextual ambidexterity. Structural ambidexterity means managing the trade-offs by implementing "dual structures" with units focussing on either exploration or exploitation. Contextual ambidexterity refers to balancing exploration and exploitation tasks within existing units, teams or individuals. DILs are organisational units bundling necessary capabilities for the development of digital innovation. The author conducted an intensive single case study on an existing DIL and provides structured insights into this organisational design and its possible contribution towards enabling organisational ambidexterity. Based on the results of the case study the author concludes that existing ambidexterity concepts do not fully describe the type of ambidexterity enabled through digital innovation labs. This organisational concept allows the full focus of certain employees on either exploration or exploitation tasks based on a temporal aspect. We follow the conclusion of the author who identifies DILs as a promising organisational design for the identification of technology potentials within organisations.

References

1. Hütter, A., Riedl, R.: Chief Information Officer Role Effectiveness: Literature Review and Implications for Research and Practice. In: Hütter, A. and Riedl, R. (eds.) Chief Information Officer Role Effectiveness: Literature Review and Implications for Research and Practice. pp. 1–30. Springer International Publishing, Cham (2017).

2. Horlacher, A., Hess, T.: What does a Chief Digital Officer do? Managerial tasks and roles of a new C-level position in the context of digital transformation. In: Proceedings of the 49th Hawaii International Conference on System Sciences. pp. 5126–5135. IEEE Computer Society, Waikoloa Village, Hawaii (2016).
3. Nan, N., Tanriverdi, H.: Unifying the Role of IT in Hyperturbulence and Competitive Advantage Via a Multilevel Perspective of IS Strategy. MIS Q. 41, 937–958 (2017).
4. Overby, E., Bharadwaj, A., Sambamurthy, V.: Enterprise agility and the enabling role of information technology. Eur. J. Inf. Syst. 15, 120–131 (2006).
5. Kahre, C., Hoffmann, D., Ahlemann, F.: Beyond Business-IT Alignment - Digital Business Strategies as a Paradigmatic Shift: A Review and Research Agenda. In: Proceedings of the 50th Hawaii International Conference on System Sciences. pp. 4706–4718. AIS Electronic Library, Waikoloa Village, Hawaii (2017).
6. Henderson, C., Venkatraman, N.: Strategic alignment: Leveraging information technology for transforming organizations. IBM Syst. J. 32, 472–487 (1999).
7. Heinrich, L.J., Riedl, R., Stelzer, D., Sikora, H.: Informationsmanagement: Grundlagen, Aufgaben, Methoden. De Gruyter Oldenbourg, München (2014).
8. Krcmar, H.: Informationsmanagement. Springer Gabler, Berlin Heidelberg (2015).
9. Riedl, R., Benlian, A., Hess, T., Stelzer, D., Sikora, H.: On the Relationship Between Information Management and Digitalization. Bus. Inf. Syst. Eng. 59, 475–482 (2017).
10. Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M.: Embracing digital technology: A new strategic imperative. MIT Sloan Manag. Rev. 55, 1–16 (2014).

Control of Systemic Risks in Global Networks – A Grand Challenge to Information Systems Research

Fabian Lorig[1], Ingo J. Timm[1], and Peter Mertens[2]

[1] Trier University, Center for Informatics Research and Technology, Trier, Germany
{lorigf,itimm}@uni-trier.de
[2] Friedrich-Alexander-University Erlangen-Nürnberg,
Wirtschaftsinformatik I, Nürnberg, Germany
peter.mertens@fau.de

Abstract. The emergence of global networks also results in the occurrence of systemic risks that might affect the stability of the overall system. To cope with these risks, this workshop on the "Control of Systemic Risks in Global Networks" provides a platform for the collection and discussion of innovative approaches, methods, and theories but also of practical problems from the areas of simulation, artificial intelligence, operations research, and statistics. This enables the exchange of experiences and methods between scientists and practitioners.

Keywords: Grand Challenge, Systemic Risk, Reference Framework.

1 Introduction

Modern communication networks lead to a stronger coupling of and interdependency between social and economic areas. Examples are electronic marketplaces, which enable ever faster transactions, worldwide production networks, which allow for higher specialization with increasing efficiency, and *smart grids*, which facilitate the provision of energy in the European Single Market by means of flexible control. The resulting worldwide and interconnected networks increasingly decide on the competitiveness of enterprises.

On the one hand, this development is promoted by a strong **demand pull** for innovative technologies that emanates from companies. This results from the companies' endeavor to take advantage of environmental differences in a "globalized world". Examples are increasing sales opportunities in emerging countries, low labor costs, special competences in the development and production of electronic components or software products, discoveries of raw materials, and tax conditions.

On the other hand, there is an increasing **technology pressure**. This is due to an increasing performance-cost ratio of data management as well as from the fact that modern multi and manycore systems accelerate or initially enable the solving of sophisticated planning, disposition, and control algorithms. Moreover, the advancement of traditional methods, e.g., artificial neural networks and deep learning,

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

allows for the discovery of patterns and the investigation of systems that remained hidden or were inaccessible before.

Along with these worldwide networks, *systemic risks* emerge which affect the stability of the overall system [1]. Examples of potential failures are flash crashes in high-frequency trading, production downtime due to delivery delays, or blackouts in energy networks. For instance, on September 28th, 2003, power plant failures in Italy lead to disruptions of the Internet infrastructure, which relied on energy supply and at the same time was required to control other power plants. This resulted in a cascade of failures and has nearly caused the collapse of the entire Italian energy supply [2,3].

Obviously, not all risks are equivalent with respect to their probability of occurrence and of the consequences. Thus, those *systemic risks* must be identified, which – as illustrated by the example – affect the stability of the overall system and are not considered as part of the risk assessment of the independent subsystems. Here, the extent of the risk must be considered as well as the probability of finding an adequate countermeasure with reasonable effort.

In a joint initiative, which is steered by the *German Informatics Society* (Gesellschaft für Informatik; GI), Information Systems Research and Computer Science have selected the *control of systemic risks in global networks* as one of the five most important Grand Challenges for the future [4]. From an information system research perspective, two major interests can be identified: On the one hand, the availability as well as the situational aggregation and interpretation of decision-relevant information and on the other hand the autonomous identification, quantitative estimation, and flexible reaction to risks.

2 Current Technology Pressure

In information system as well as computer science research, there are ongoing discussions whether networks can be designed or dynamically emerge from the interaction of devices with network technologies: Worldwide networks are not designed as part of an "engineering process", they are created through the interaction of interconnected systems as emergent phenomenon and must be described and understood [5].

The need for a development of methods for the design of such networks can be identified when investigating the current technology pressures. Developments that can contribute to the control of systemic risks include but are not limited to:

1. **Communication Networks**: Advances in communication networks, e.g., an increasing performance-cost ratio of communication channels (hardware) and greater flexibility in routing (software), which allow for prioritized communication in case of emergency.
2. **Simulation**: Recent developments in simulation from a tool for planning support to a real-time assistance for decision support through the development of innovative formalisms, e.g., system dynamics or agent-based simulation, and due to the immediate availability of current data.

3. **Machine Learning**: Revolutionary progress in machine learning that is facilitated by the increasing availability and amount of (training) data as well as shift from multi to multi and manycore computing. This allows for the use of deep learning, convolutional neural networks as well as data, text, and opinion mining techniques.
4. **Decentralized Control**: The availability of approaches for decentralized and adaptive control with autonomous software agents, multiagent systems, and organic computing promotes the high-tech strategy "Industry 4.0".
5. **Transaction Processing Systems (Blockchain)**: New forms of transaction processing systems, e.g., blockchain, allow for the tamper-resistant and decentralized organization and logging of safety-critical operations in processes such as access or updates of sensitive data.
6. **Multilayer and Multiplex Networks**: A shift from the analysis of isolated and homogenous networks to the investigation of multilayer and multiplex networks (interdependent networks).
7. **Convergence**: The convergence of technical systems and processes leads to the unification of business models and technologies across sectors. Through this, technical and economic success of one domain might dominate another domain, e.g., successful business models of internet giants can compete with stationary trade in the physical world even though the horizon of experience is considerably lower.

Due to disciplinary barriers, the aforementioned technology areas are not yet sufficiently developed, applied, or transferred for controlling systemic risks. This limits the opportunities for action that can be undertaken to prevent the potentially dramatic consequences of systemic risks. Still, these technologies have a high potential to contribute as component of a solution for controlling systemic risks.

Considering disaster management strategies, for instance, it can be illustrated how disciplines can learn from each other and benefit from the experiences of other disciplines. Insurance companies make use of reinsurances to handle major claims which could result in their insolvency. Such approaches are also applicable to supply chain management as protection against supply shortages that might result in disruptions of the own production of goods. In this regard, supply chain management can also learn from insurances as systemic risks emerge from networks of reinsurances which can potentially result in uncontrollable chain effects that lead to global crises.

3 Reference Framework

Suitable technologies and methods for controlling systemic risks are diverse. Thus, to classify and distinguish different approaches, we suggest the use of a morphological box. It serves as a reference framework for discussion within the workshop as approaches can be classified and assessed according to different dimensions. In Figure 1, the morphological box is illustrated that is used for the assessment of the approaches that are presented as part of this workshop. For each approach, the aspects of *networks*, *risks*, and *decision situation* are focused.

To this end, the domain focus of the workshop lies on *logistics*, *finance & insurances*, and *public services*, yet, also contributions from other domains are welcome. With respect to the type of risk that is addressed by the approaches, it can be differentiated into five types, according to the domain the risk is related to: *production*, *market*, *finance*, *institution*, and *nature*. In addition, also the occurrence of the risk is classified as *regularly*, *periodically*, or *rarely*. Finally, the decision situation of the risk can be specified according to the risk's predictability as well as by the authority which is the decision maker.

Network	***Domain***	Logistics	Finance & Insurances	Public services		
	Network model	Available and fixed	Available and ongoing change	Situation-dependent change	Ad-hoc	
Risk	***Type of risk***	Production	Market	Finance	Institutional	Nature
	Occurrence	Regularly	Periodically	Rarely		
Decision situation	***Predictability***	Predictable and plannable	Predictable and not plannable	Not predictable but plannable	Not predictable and not plannable	
	Authority	Single person	Committee	Automated, with intervention of persons	Fully automated	
	Horizon	Enterprise	Network	Society		

Figure 1: Reference framework for the classification and discussion of approaches.

4 Discussion

To address the Grand Challenge of controlling systemic risks in global networks, this workshop aims at both the collection and discussion of innovative approaches, methods, and theories but also practical problems from the areas of simulation, artificial intelligence, operations research, and statistics. To this end, the goal of the workshop is to provide a platform for the exchange of experiences and methods between scientists and practitioners. Moreover, the development of a medium-term research agenda shall be promoted for targeting this Grand Challenge.

5 Acknowledgements

We would like to thank all those that contributed to the initiative on "Control of Systemic Risks in Global Networks" for their fruitful comments and innovative ideas.

References

1. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. In: Nature, 464 (p. 15).

2. UCTE (2004). FINAL REPORT of the Investigation Committee on the 28 September 2003 Blackout in Italy. UCTE Report – April 2004.
3. Johnson, C. W. (2007). Analysing the causes of the Italian and Swiss blackout, 28th September 2003. In *Proceedings of the twelfth Australian workshop on Safety critical systems and software and safety-related programmable systems. Volume 86* (pp. 21-30). Australian Computer Society, Inc.
4. Mertens, P. & Barbian, D. Business & Information Systems Engineering (2015) 57: 391.
5. Gkantsidis, C., Mihail, M., & Saberi, A. (2003). Conductance and congestion in power law graphs. In: ACM SIGMET-RICS Performance Evaluation Review Vol. 31, No. 1 (pp. 148-159), ACM.

Die Mitarbeiter von morgen - Kompetenzen künftiger Mitarbeiter im Bereich Business Analytics

Uwe Neuhaus[1], Hinrich Schröder[1], and Michael Schulz[1]

[1] NORDAKADEMIE Hochschule der Wirtschaft, Fachbereich Informatik, Elmshorn, Germany
{uwe.neuhaus,hinrich.schroeder,michael.schulz}@nordakademie.de

Abstract. Im hier beschriebenen Workshop soll diskutiert werden, welche Kompetenzen die „Mitarbeiter von morgen" in der digitalen Transformation, speziell im Bereich Business Analytics, aufweisen müssen und wie diese erlangt werden können. Workshopteilnehmer vertreten jeweils eine der folgenden drei Gruppen: (1) Unternehmen, die unter dem Druck des Fachkräftemangels und dem steigenden Grad der Digitalisierung auf der Suche nach geeignetem Personal sind, (2) aktuell oder zukünftig im Umfeld Business Analytics tätige Personen, die eine den Anforderungen entsprechende Ausbildung benötigen, sowie (3) Lehrstätten, also allgemein- und berufsbildende Schulen, Weiterbildungseinrichtungen und vor allem Hochschulen, die junge Menschen auf die Arbeitswelt von morgen vorbereiten sollen. Als Arbeitsziel ist das Diskutieren von gemeinsamen und unterschiedlichen Sichtweisen, sowie das Ableiten von Handlungsempfehlungen zu sehen.

Keywords: Digitale Transformation; Business Analytics; Mitarbeiter; Kompetenzen

1 Mitarbeiter als Erfolgsfaktor der digitalen Transformation

Digitalisierungsvorhaben in Unternehmen weisen unterschiedliche Zielrichtungen auf: Produkte und Leistungsangebote des Unternehmens können verändert, digitale Services kreiert sowie alle Phasen der Kundenbeziehung völlig neu gestaltet werden [1]. In einer eher nach innen gerichteten Betrachtung steht die Digitalisierung der Unternehmensprozesse im Fokus. Neben unterschiedlichen technologischen Treibern wie Cloud Computing, Robotik oder Künstliche Intelligenz bildet die Sammlung, Aufbereitung und Analyse von Daten in vielen dieser Vorhaben eine wichtige Grundlage.
Zur Umsetzung von Digitalisierungsvorhaben sind Voraussetzungen zu schaffen, die im Regelfall zu simultanen und tiefgreifenden Veränderungen in Unternehmen führen und den vielzitierten Begriff der „digitalen Transformation" prägen (zum Begriff vgl. [2]). Die notwendigen Veränderungen lassen sich den Bausteinen Technologie, Führung und Kultur, Struktur und Organisation, Kollaboration mit Partnern, Prozessarchitektur sowie Kompetenzen von Mitarbeitern zuordnen [1]. Müller et al. [3] fassen diese Zusammenhänge in einem Steuerungsmodell für die digitale Transformation zusammen (vgl. Figure 1).

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

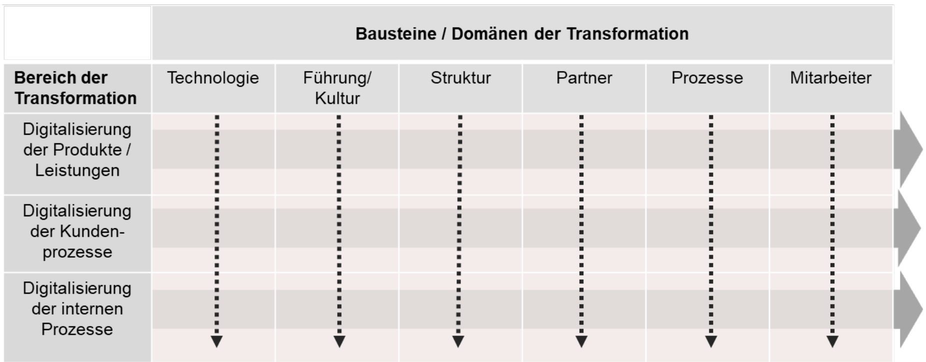


Figure 1. Transformations-Matrix (in Anlehnung an [3])

Besonders der letztgenannte Baustein – die Veränderungen auf Ebene der Mitarbeiter – kann als zentraler Erfolgsfaktor einer Transformation angesehen werden.
Die Akquisition und Entwicklung qualifizierter Mitarbeitern mit den für die digitale Transformation benötigten Kompetenzen stellt für viele Unternehmen angesichts demografischer Entwicklungen und massiver Veränderungen der Anforderungsprofile eine große Herausforderung dar [4].
In dem Workshop sollen diese Überlegungen aufgegriffen und die Frage diskutiert werden, welche Kompetenzen die „Mitarbeiter von morgen" in der digitalen Transformation aufweisen müssen und wie diese erlangt werden können.
Auch wenn Auswirkungen der Digitalisierung auf die Arbeit für nahezu alle berufstätigen Personen spürbar sind, vollzieht sich die Veränderung nicht überall im gleichen Maß. Die spezielle Betrachtung des Bereiches Business Analytics in diesem Workshop ist damit zu rechtfertigen, dass hier eine besonders starke Veränderung des Arbeitsumfeldes zu beobachten ist: Unternehmen haben in den letzten Jahren verstärkt in den Ausbau ihrer Business-Analytics-Systeme investiert. Dies liegt neben der grundlegenden Entwicklung durch die Digitalisierung daran, dass technische Restriktionen, die Analysen in der Vergangenheit langwierig und teuer gemacht haben, nur noch eine untergeordnete Rolle spielen [5]. Außerdem liegen immer mehr Daten heterogener Struktur, deren Analyse einen Mehrwert generiert, in digitaler Form vor [6] und nicht zuletzt ist verstärkt die Erkenntnis aufgekommen, dass neben strategischen auch operative Entscheidungen möglichst datenbasiert getroffen werden sollten [7].
Um die aus den genannten Entwicklungen resultierenden heterogener gewordenen Analyseanforderungen verschiedener Domänenexperten[1] in angemessener Zeit befriedigen zu können, wurden in den letzten Jahren einfache Softwareprodukte entwickelt, die es auch Gelegenheitsanwendern in Unternehmen ermöglichen sollen, (teilweise) eigenständig Standardberichte aufzubauen oder Ad-hoc-Analysen durchzuführen – ohne dabei Business-Analytics-Experten einbeziehen zu müssen [8]. Dies führt durch eine zeitgleich stattfindende Entwicklung jedoch nicht zu einer Substitution letzterer,

[1] Der Begriff Domänenexperte wird hier als Abgrenzung zu einem Business-Analytics-Experten verwendet.

sondern zu einer Verschiebung der Tätigkeitsfelder, die in vielen Berufen als eine Folge der digitalen Transformation zu beobachten ist [9]. Im hier fokussierten Umfeld schlägt sich diese Verschiebung durch einen vermehrten Einsatz fortgeschrittener Analysemethoden – häufig mit Begriffen wie Data Mining, Data Science oder maschinelles Lernen betitelt – nieder.

Die beschriebenen Veränderungen haben zur Folge, dass die Kompetenzen im Umfeld analytischer Informationssysteme, die bisher in Lehrstätten vermittelt wurden, nicht mehr geeignet sind, um sämtliche, an die Mitarbeiter herangetragenen Aufgaben lösen zu können. Generell können Mitarbeiter, deren Berufsfeld stark von den Auswirkungen der Digitalisierung betroffen ist, nicht davon ausgehen, dass die über Ausbildung und Erfahrung gewonnenen Kompetenzen ausreichen, die langfristig an sie gestellten Anforderungen zu bewältigen ([10 - 11]). Zur Sicherstellung einer dauerhaften Erwerbstätigkeit ist eine kontinuierliche Aus- und Weiterbildung vonnöten [12]. Weiterhin ist eine Individualisierung der Ausbildung erforderlich, die für die Mitarbeiter und das Arbeitsumfeld relevante Kompetenzen vermittelt [13].

Um geeignete Anforderungen an die Mitarbeiter von morgen und den damit verbundenen Veränderungsbedarf zu identifizieren sind drei Interessengruppen zu berücksichtigen:

(1) Unternehmen, die unter dem Druck des Fachkräftemangels und dem steigenden Grad der Digitalisierung auf der Suche nach geeignetem Personal sind,
(2) aktuell oder zukünftig im Umfeld Business Analytics tätige Personen, die eine den Anforderungen entsprechende Ausbildung benötigen, sowie
(3) Lehrstätten, also allgemein- und berufsbildende Schulen, Weiterbildungseinrichtungen und vor allem Hochschulen, die junge Menschen auf die Arbeitswelt von morgen vorbereiten sollen.

2 Mitarbeiter von heute

Ausgangspunkt für die Bestimmung der Anforderungen an die Mitarbeiter von morgen ist zunächst die Erfassung des Status quo. Dies bedingt die Beantwortung der Frage, welche Kenntnisse und Qualifikationen Personen, die im Bereich Business Analytics tätig sein wollen, bereits heute mitbringen sollten. Neuhaus & Schulz analysierten im November 2017 500 Stellenanzeigen im Kontext analytischer Informationssysteme [14]. Ziel dieser Untersuchung war es zu bestimmen, welche konkrete Bedeutung den häufig verwendeten Schlagwörtern *Analytics*, *Business Intelligence*, *Data Analytics*, *Data Science*, *Advanced Analytics* und *Business Analytics* in der betrieblichen Praxis gegeben wird. Durch die eingesetzte Untersuchungsmethode (Frequenzanalyse der in den Stellenanzeigen verwendeten Begriffe) entstand gleichzeitig aber auch ein präzises Anforderungsprofil der gesuchten Business-Analytics-Experten. Außerdem wurde deutlich, welche Tätigkeiten sie ausführen sollen, in welchen Branchen sie besonders gefragt sind und welche treibenden Kräfte ihr Arbeitsumfeld beeinflussen. Figure 2 gibt einen Überblick über die identifizierten Dimensionen und ihre Kategorien.

Table 1. Verzeichnis verwendeter Abkürzungen

Aufgabentypen	ANA	Analyse
	BEW	Bewertung
	MONI	Monitoring
	PLAN	Planung
	PROG	Prognose
	STEU	Steuerung
	VISU	Visualisierung
Analyse-methoden	ADHOC	Ad-hoc-Analysen
	DMKL	Klassische Data Mining-Methoden
	REP	Reporting
	SPEZM	Spezielle Methoden
	STAT	Statistische Methoden
Anwendungs-felder	KUND	Kundenanalysen
	MARKT	Marktanalysen
	MKT	Marketinganalysen
	RSK	Risikoanalysen
	SPEZ	Spezialisierte Analysen
	UNTW	Weitere Unternehmensanalysen
	WEB	Analysen von Websites (eigene und fremde)
Branchen	BERA	Beratung
	DIENS	Weitere Dienstleistungen
	FINA	Finanzen
	INDU	Industrie
	MARK	Marketing
Datenma-nagement	DAUFB	Datenaufbereitung
	NTDB	Nicht-traditionelle Datenbankmanagementsysteme
	STRDA	Strukturierte Daten
	SUDA	Semi- und unstrukturierte Daten
	TDB	Traditionelle Datenbankmanagementsysteme
Formale Qualifikatio-	INF	Informatik
	ING	Ingenieurswesen
	MATH	Mathematik
	NAWI	Naturwissenschaften
	WINF	Wirtschaftsinformatik
	WIRTS	Wirtschaft
Program-mierung	DASP	Datenanalysesprachen
	DBSP	Datenbanksprachen
	DOSP	Andere domänenspezifische Sprachen
	UNIVP	Universelle Programmiersprachen
	WEBS	Web-Sprachen
Treiber	BIGD	Big Data
	DIGIT	Digitalisierung
	IND4	Industrie 4.0
	INTDD	Internet der Dinge
	REGU	Regulatorik

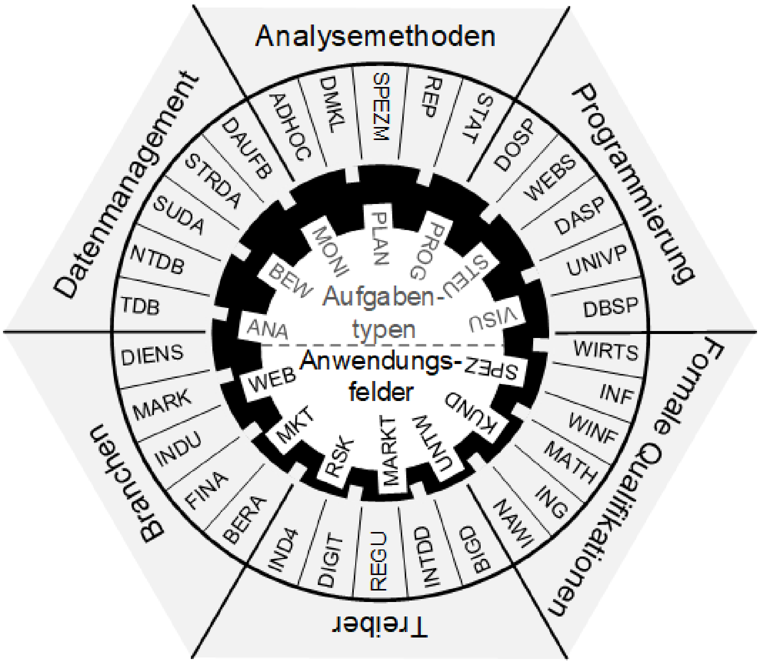


Figure 2. Relevante Dimensionen analytischer Informationssysteme und deren Kategorien (nach [14])

Aufgabentypen und Anwendungsfelder
Analytische Fähigkeiten stehen naturgemäß ganz oben auf der Liste der gesuchten Qualifikationen, meist jedoch in Zusammenhang mit ganz spezifischen Anwendungsfeldern. Hier sind insbesondere die *Analyse von Webdaten* und von *Marketingmaßnahmen* zu nennen, mit etwas Abstand folgen *Markt-*, *Kunden-* und *Risikoanalysen*. Häufig bildet die Analyse auch nur den ersten Arbeitsschritt, dem dann weitere Schritte wie *Steuerung*, *Prognose*, *Planung* und *Bewertung* folgen sollen. Von Bedeutung sind ebenfalls die *Visualisierung* von Analyseergebnissen sowie das *Monitoring* von Systemen.

Treiber
Die Begriffe *Digitalisierung* und *digitale Transformation* werden in den untersuchten Stellenanzeigen erwartungsgemäß häufig als Treiber für die Suche nach Mitarbeitern im Bereich analytischer Informationssysteme angegeben. Häufiger allerdings noch wird *Big Data* genannt. Da in den Stellenanzeigen jedoch selten konkretes Big-Data-Know-how nachgefragt wird, ist allerdings zu vermuten, dass dieser Begriff offenbar auch verwendet wird, wenn es generell um Datenanalyse geht und nicht speziell um die Analyse von großen, komplexen oder sich schnell ändernden Datenmengen. Ähnlich häufig wie Digitalisierung fanden sich Treiber aus dem Themenbereich *Regulatorik*. Typische Schlagwörter sind hier etwa Compliance, Audit und Governance. Stärker technisch orientierte Treiber, die noch nennenswert häufig erwähnt wurden, sind *Industrie 4.0* und *Internet der Dinge*.

Branchen
In zwei Branchen sind Experten für analytische Informationssysteme besonders stark nachgefragt: in der *Beratungs*- und in der *Finanzbranche*. Beides verwundert nicht. Beratungsunternehmen sind häufig Vorreiter bei der Einführung innovativer Technologien und häufig bedienen sich Unternehmen, die selbst noch nicht über ausreichendes Technologie-Knowhow verfügen, der Dienste von Beratern. In der Finanzindustrie spielen Datenanalysen traditionell eine große Rolle, sodass die Nutzung innovativer Analysemethoden naheliegt. Außerdem unterliegen Banken und Versicherungen vielfältigen Regularien, was – wie oben erwähnt – ein weiterer Treiber für den Einsatz analytischer Informationssystem ist. Weitere Branchen, die überdurchschnittlich viele Analyseexperten suchen, sind die *Industrie* (insbesondere der Maschinenbau, die Elektro- und Automobilindustrie) im Kontext der Treiber Industrie 4.0 und Internet der Dinge sowie die *Marketingbranche* mit ihren zunehmend wichtiger werdenden Datenquellen aus den Bereichen Online-Marketing, Mobile-Marketing und Social Media. Eine nennenswerte Anzahl an Stellenanzeigen finden sich außerdem noch in den Bereichen *Logistik*, *Gesundheit* und *Medien*.

Formale Qualifikationen
Betrachtet man die formalen Qualifikationen, die von den zukünftigen Mitarbeitern im Bereich analytischer Informationssysteme erwartet wird, so fällt zunächst auf, dass fast ausschließlich Akademiker gesucht werden, nur vereinzelt werden auch Bewerber mit einem Ausbildungsberuf angesprochen. Dies deutet auf die Komplexität der zu übernehmenden Aufgaben hin. Explizit genannt werden in den Stellenanzeigen außerdem fast ausschließlich traditionelle Studienabschlüsse. Spezialisierte Studienabschlüsse wie etwa als ‚Data Scientist' werden von den Bewerbern (noch) nicht gefordert. Am häufigsten gesucht werden Absolventen aus den Bereichen *Wirtschaftswissenschaften/BWL* und *Informatik*. Auch *Wirtschaftsinformatiker*, die ein Bindeglied zwischen diesen beiden Disziplinen darstellen, sind stark nachgefragt. In stärker technisch orientierten Branchen oder für spezielle Analyseformen werden aber auch *Mathematiker*, *Ingenieure* und *Naturwissenschaftler* angesprochen.

Datenmanagement
Die in den Stellenanzeigen explizit gesuchten fachlichen Fertigkeiten lassen sich grob in die drei Dimensionen Datenmanagement, Analysemethoden und Programmierung aufteilen. Das *Datenmanagement* umfasst nach Gandomi und Haider [15] Prozesse und Technologien, um Daten zu beziehen, zu speichern und für die Analyse vorzubereiten. Die zentrale Fertigkeit innerhalb des Datenmanagements ist die in 70 % aller Stellenanzeigen geforderte Kompetenz zur *Datenaufbereitung*. Erwartet werden hier von den Bewerbern Kenntnisse sowohl im Umgang mit *strukturierten* als auch mit *unstrukturierten Daten*. Dies deckt sich mit den Aussagen von Russom [16] und Chamoni [17], welche die zunehmende Bedeutung von semi- und unstrukturierten Daten für analytische Informationssysteme hervorheben. Als Datenquelle von großer Bedeutung sind beim Datenmanagement weiterhin Datenbanken. Sowohl Kenntnisse relationaler Datenbankmanagementsysteme (DBMS) als auch nicht-traditioneller DBMS (z. B. In-

Memory-Datenbanken, No-SQL-DBMS, multidimensionale DBMS) werden oft gesucht. Häufig wird in den Stellenanzeigen dabei Know-how in spezifischen Software-Produkten erwartet. Die Menge und Vielfalt der genannten Produkte ist dabei sehr groß, kein einzelner Hersteller nimmt eine dominierende Marktstellung ein.

Analysemethoden

Die Mehrzahl der untersuchten Stellenanzeigen erwartet von den gesuchten Experten auch die Beherrschung verschiedener *Analysemethoden*. Im Mittel werden in den Anzeigen drei bis vier solcher Methoden aufgezählt. Fast alle Methoden lassen sich fünf verschiedenen Kategorien zuordnen. Der am häufigsten genannte Begriff ist dabei das *Reporting*, die Aufbereitung der Analyseergebnisse in eine verständliche Darstellungsform. Hier scheint insbesondere die Datenaufbereitung in Form von Dashboards an Bedeutung zu gewinnen. Das Reporting ergänzt die *Ad-hoc-Analyse*, welche Entscheidungsträgern die unmittelbare, spontane Untersuchung von Analyseergebnissen ermöglicht. Wichtige Begriffe in diesem Zusammenhang Online Analytical Processing (OLAP), Self-Service-BI und Visual Analytics. Viele nachgefragte Analysemethoden entstammen auch der *Statistik*. Dazu zählen etwa die Regression, multivariate Statistik und AB-Tests. Die letzten beiden Analysemethodenkategorien lassen sich beide dem Bereich Data-Mining zuordnen. *Klassische Data-Mining-Methoden* stellen grundlegende Verfahren dar, die im Data-Mining weit verbreitet sind und bei vielen unterschiedlichen Analyseproblemen genutzt werden können. Beispiele sind etwa Clusteranalysen, Entscheidungsbäume oder Assoziationsanalysen. *Spezielle Data-Mining-Methoden* erfordern hingegen spezielles Experten-Know-how oder lassen sich nur auf spezifische Analysefragestellungen anwenden. Hierzu zählen z. B. Deep-Learning, Big-Data-Analytics oder Text-Mining. Wie schon beim Datenmanagement werden auch bei den Analysemethoden häufig Kenntnisse spezieller Software-Produkte gesucht. Die Vielfalt der Produkte ist hier sogar noch größer als im Datenmanagementbereich. Neben den Lösungen großer Software-Hersteller finden sich noch zahlreiche Produkte spezialisierter Anbieter sowie Open-Source-Werkzeuge.

Programmierung

Da auch bei Nutzung von Datenanalyse-Software häufig die Notwendigkeit entsteht, bestimmte Schritte des Analyseprozesses durch selbstgeschriebene Routinen umzusetzen, werden von Experten analytischer Informationssysteme häufig auch Kenntnisse in *Programmierung* erwartet. Auch hier lassen sich fünf Hauptkategorien identifizieren. Die ersten vier Kategorien umfassen dabei domänenspezifische Sprachen, also Sprachen die für einen speziellen Einsatzzweck entworfen wurden. In Ergänzung zum Datenmanagement sind häufig *Datenbanksprachen* von Bedeutung, insbesondere SQL und seine herstellerspezifischen Erweiterungen wie PL/SQL oder Transact-SQL. Die zweite Kategorie bilden *Web-Sprachen* wie JavaScript und PHP, die häufig zusammen mit der Auszeichnungssprache HTML genannt werden. Ihre Bedeutung erhalten diese Sprachen durch den hohen Bedarf im Bereich Web-Analytics sowie der zunehmenden Nutzung von web-basierten Analyse-Dashboards. *Datenanalysesprachen* wurden speziell für die Bearbeitung von analytischen Fragestellungen entworfen. Sehr dominant war in den untersuchten Stellenanzeigen die für statistische Datenanalyen entwickelte

Programmiersprache R. *Weitere domänenspezifische Sprachen* wie ABAP oder VBA verdanken ihre Nachfrage der weiten Verbreitung spezieller, herstellerspezifischer Systeme (SAP-Systeme bzw. Microsoft-Office-Produkte). *Universelle Programmiersprachen* wurden im Gegensatz zu domänenspezifischen Sprachen zur Software-Entwicklung in beliebigen Anwendungsbereichen entworfen. Daher können sie auch im Kontext analytischer Informationssysteme genutzt werden. Die am stärksten nachgefragten universellen Programmiersprachen waren in den Stellenanzeigen mit deutlichem Vorsprung Python (53 %), gefolgt von Java (20 %).

3 Kompetenzanforderungen an den Mitarbeiter von morgen

Auf Basis der in Kapitel 1 genannten Entwicklungen und der Erkenntnisse aus der in Kapitel 2 beschriebenen Ist-Analyse können Anforderungen an die Mitarbeiter von morgen im Bereich Business Analytics für mindestens drei Gruppen diskutiert werden:

(1) Mitarbeiter, die im Bereich Business Analytics tätig sind und vor der Herausforderung stehen, Kompetenzen aufbauen zu müssen, um den neuen Anforderungen an ihr Berufsfeld zu genügen. (Data Scientists und Business-Analytics-Experten)

(2) Mitarbeiter, deren Kerntätigkeit nicht im Bereich der Business Analytics liegt, die jedoch vor der neuen Anforderung stehen, eigenständig, einfache Analysen durchzuführen. Zudem wird von ihnen in einem gewissen Maße die Fähigkeit erwartet, auch Ergebnisse fortgeschrittener Analysen verstehen, bewerten und einordnen zu können. (Domänenexperten)

(3) Mitarbeiter, die durch geeignete Kompetenzen in der Lage sind, neue datenbasierte Produkte, Services oder Prozesse zu entwickeln und damit die zukünftige Wettbewerbsfähigkeit eines Unternehmens sicherstellen können. Sie müssen zusätzlich die Kompetenzen besitzen, die dafür notwendigen Veränderungsprozesse in Unternehmen umzusetzen und in der Organisationshierarchie entsprechend positioniert sein. (Entscheider)

Für alle Mitarbeitergruppen und auch für die Unternehmen, in denen die Mitarbeiter tätig sind, erwachsen Herausforderungen aus den zusätzlichen, sich schnell wandelnden Aufgabenbereichen [10], die im Folgenden einzeln diskutiert werden.

Data Scientists und Business-Analytics-Experten

Während für Business-Analytics-Experten traditionell das Standard- und Ad-hoc-Reporting, der Prozess der Datenaufbereitung und die geeignete Datenhaltung im Fokus ihrer Aufgaben standen, verlangt die Anwendung fortgeschrittener Analysemethode nach mathematischen Fähigkeiten, insbesondere im Bereich der Statistik, nach Kenntnissen in speziellen Programmiersprachen und nach einem Verständnis von Domänen, das weit über den traditionellen Anwendungsbereich analytischer Informationssysteme hinausgeht [14]. Um den Anforderungen gerecht zu werden, hält es die Mehrzahl von

in diesem Bereich tätigen Personen für notwendig, zusätzliche Methodenkompetenzen aufzubauen [18]. Auch eine stärkere Domänenfokussierung bei der Identifikation für den Aufgabenbereich benötigter Kompetenzen kann relevant sein: Während in der Vergangenheit primär strategisch-betriebswirtschaftliche Fragestellungen im Fokus analytischer Informationssysteme standen, sind es nun zusätzlich auch ganz andere, wie etwa aus dem Bereich der Operational Business Intelligence [19] oder der Predictive Maintenance [20]. Zusammenfassend kann weder aus Domänen- noch aus Methodensicht davon ausgegangen werden, dass ein Mitarbeiter in der erweiterten Ausrichtung des Aufgabenfeldes in der Lage ist, bei sämtlichen Unternehmensfragestellungen unterstützen zu können.
Der Wegfall traditioneller Tätigkeiten von Business-Analytics-Experten führt zu einer Reduzierung der Routineaufgaben, was ein Arbeitsumfeld auf der einen Seite zwar attraktiver, auf der anderen Seite aber auch anspruchsvoller macht [21]. Ein daraus resultierender Bedarf der Übernahme zusätzlicher Aufgaben führt wie beschrieben auch zu einem zusätzlichen Qualifizierungsbedarf. Dies stellt Mitarbeiter vor die Herausforderung der Auswahl geeigneter Angebote. Während das Ausbildungsangebot bereits sehr intransparent ist, zeigt sich der Markt der Weiterbildungsangebote noch unübersichtlicher [22]. Die Entscheidung für eine ungeeignete Entwicklung der eigenen Kompetenzen birgt eine Gefahr für die langfristige Sicherstellung der Beschäftigung.

Domänenexperten
In der Vergangenheit waren Domänenexperten vor allem Konsumenten von Standardberichten oder konnten gegebenenfalls Ad-hoc-Analysen auf Basis einfacher Datenmodelle durchführen. Heute benötigt jeder Mitarbeiter eine Basisausbildung im Bereich der Digitalisierung, die schon in der Regelschule beginnen [12] und in weitergehenden Ausbildungen ausgebaut werden muss. Im Kontext der Business Analytics kann diese vor allem in der Datenkompetenz, also der Fähigkeit Daten als Grundlage der eigenen Arbeit geeignet einsetzen zu können [23], gesehen werden. Anders als häufig angenommen, kann nicht davon ausgegangen werden, dass Digital Natives nur deshalb eine höhere Datenkompetenz besitzen, weil sie von Kind auf mit digitalen Produkten vertraut sind [24].

Entscheider
Die Entscheidung zur Digitalisierung von Geschäftsprozessen muss häufig unter großer Unsicherheit getroffen werden [10] und hat häufig umfassende Auswirkungen auf Unternehmen [25]. Trotzdem fehlt es vielen Organisationen an einer klaren Strategie, wie mit dem Thema der Transformation umgegangen werden soll [26]. Mitarbeiter, die die zukünftige Wettbewerbsfähigkeit eines Unternehmens sicherstellen müssen, benötigen nicht nur Methoden- und Domänen-, sondern zusätzlich auch Veränderungs- und Gestaltungskompetenz [10] und die Fähigkeit bzw. Stellung im Unternehmen notwendige Änderungen durchzusetzen, ohne das unternehmerische Handeln außer Acht zu lassen [27].

Neben den oben genannten, stark auf den Business-Analytics-Komplex fokussierten Aspekten sind bei der Diskussion der Anforderungen an den Mitarbeiter von morgen auch immer Themen der digitalen Transformation zu berücksichtigen, die das Arbeitsleben an sich verändern. Genannt werden können hier etwa die durch die Verwendung mobiler Endgeräte ermöglichte Flexibilisierung von Arbeitszeit und –ort, sowie die teilweise Aufhebung der Grenzen zwischen Arbeits- und Freizeit [28 – 29]. Auch ist nicht abzusehen, inwieweit die Digitale Transformation die Unterscheidung nach typischen Männer- und Frauenberufen aufzubrechen vermag, um eine geschlechterübergreifende Chancengleichheit auf dem Arbeitsmarkt auch im Bereich informationstechnisch fokussierter Berufe herzustellen [30].

4 Ziel des Workshops

Die Leitfragen die dem Workshop zu Grunde gelegt werden, orientieren sich an folgenden Kompetenzfeldern, die von Zschech et al. [31] übernommen und angepasst wurden:

(1) Basiskompetenzen, die unabhängig von dem individuellen Aufgabenbereich der Mitarbeitern digitaler Unternehmen erwartet werden,
(2) befähigende Kompetenzen, die systembezogene Kompetenzen genauso beinhaltet wie soziale Kompetenzen und eine Themenaffinität und
(3) vertiefende Kompetenzen mit einer Spezialisierung auf Aufgaben und Konzepte.

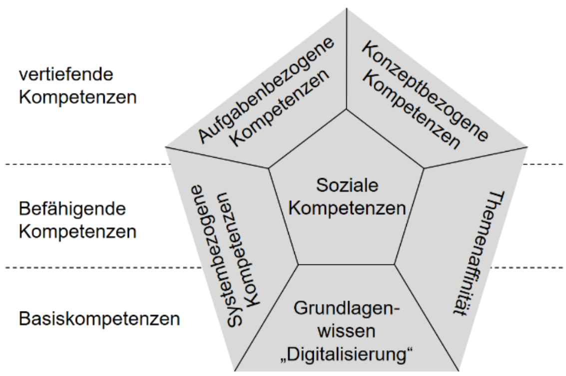


Figure 3. Kompetenzfelder der Mitarbeiter von morgen

Das Hauptaugenmerk des Workshops soll darauf gelegt werden,

- welche Kompetenzen Personen benötigen, die aktuell oder zukünftig im Umfeld der Business Analytics tätig sind,
- welche zukünftigen Herausforderungen und Verschiebungen von Aufgabenfeldern in dem Tätigkeitsumfeld gesehen werden,
- welche Erwartungen Unternehmen an die Qualifikation und Weiterentwicklung ihrer Mitarbeiter haben,

- ob der Fokus auf der Ausbildung in Softwareprodukten, Programmiersprachen oder Methoden liegen soll,
- welche Unterstützung und Informationen Mitarbeiter bezogen auf (Weiter-) Bildungsangebote seitens der Unternehmen und der Lehrstätten benötigen,
- welche Möglichkeiten existieren, um ein frühes Interesse für die Ausbildung von Business-Analytics-Kompetenzen zu entwickeln,
- wie Mitarbeiter an Lehrstätten in die Lage versetzt werden können, geeignete Lehre anzubieten,
- welche Form der Lehre geeignet ist, um benötigte Kompetenzen zu vermitteln
- und welche Konsequenzen aus den generellen Veränderungen der Arbeitswelt durch die Digitalisierung für Mitarbeiter im Business-Analytics-Umfeld zu berücksichtigen sind.

Der Zielsetzung des Workshops folgend ergibt sich aus der Beantwortung dieser Fragestellungen ein Set von Handlungsempfehlungen, das von Unternehmen und Lehrstätten genutzt werden kann, um den Herausforderungen der Ausbildung und Qualifikation der „Mitarbeiter von morgen" künftig besser begegnen zu können.

References

1. Ross J. W. et al.: Designing Digital Organizations, CISR Working Paper No. 406, Cambridge MA (2016)
2. Schallmo, D., Rusniak, A.: Roadmap zur Digitalen Transformation von Geschäftsmodellen. In: Schallmo, D. et al. (eds.) Digitale Transformation von Geschäftsmodellen – Grundlagen, Instrumente und Best Practices. Springer Gabler, Wiesbaden, pp. 1–31 (2017)
3. Müller, A., Schröder, H., von Thienen, L.: Von Entdeckern und Pionieren – Steuerungsmodell für die digitale Transformation, HMD 55: 252.; pp. 252–270 (2018)
4. Urbach, N.; Ahlemann, F.: IT-Management im Zeitalter der Digitalisierung. Auf dem Weg zur IT-Organisation der Zukunft. Springer Gabler, Berlin Heidelberg (2016)
5. van der Lans, R.: Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses. Morgan Kaufmann, Waltham (2012)
6. Larson, D., Chang, V.: A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science, in: International Journal of Information Management, 36. Jg., Heft 5, pp. 700 – 710 (2016)
7. Böhringer, M., Gluchowski, P., Kurze, C., Schieder, C.: A Business Intelligence Perspective on the Future Internet, AMCIS 2010 Paper 267, Lima (2010)
8. Alpar, P., Schulz, M.: Self-service business intelligence. Business & Information Systems Engineering, 58(2), pp. 151-155 (2016)
9. Arnold, D., Arntz, M., Gregory, T., Steffes, S., Zierahn, U.: Herausforderungen der Digitalisierung für die Zukunft der Arbeitswelt, ZEW policy brief, Nr. 8 (2016)
10. Richter, G., Ribbat, M., Thomson, B.: Digitalisierung der Arbeit und arbeitsintegriertes Lernen, in: Wulfsberg, J., Redlich, T., Moritz, M. (eds.): 1. Interdisziplinäre Konferenz zur Zukunft der Wertschöpfung pp. 83 – 92 (2016)
11. Hemkes, B., Zaviska, C.: Fachkonferenz Berufsbildung 4.0 – Zukunfschancen durch Digitalisierung, Forum 5: Berufliche Weiterbildung im Kontext von Digitalisierung – Anforderungen, Tätigkeitsverschiebungen und Gestaltungsmöglichkeiten (2017)

12. Gebhardt, J., Grimm, A., Neugebauer, L. M.: Entwicklungen 4.0 – Ausblicke auf zukünftige Anforderungen an und Auswirkungen auf Arbeit und Ausbildung. Journal of Technical Education (JOTED), Jg. 3 (Heft 2), pp. 45-61 (2015)
13. Sachs, S., Meier, C., McSorley, V.: Digitalisierung und die Zukunft kaufmännischer Berufsbilder – eine explorative Studie, Schlussbericht, Hochschule für Wirtschaft in Zürich (2016)
14. Neuhaus, U., Schulz, M.: Wie nennt ihr, was ihr da tut? Der Begriff „Analytics“ in der betrieblichen Praxis, NORDBLICK 6, pp. 4 – 29 (2018)
15. Gandomi, A., Haider, M.: Beyond the Hype – Big Data Concepts, Methods, and Analytics, in: International Journal of Information Management, 35 Jg., Heft 2, pp.137 – 144 (2015)
16. Russom, P.: BI Search and Text Analytics – New Additions to the BI Technology Stack, TDWI Best Practice Report (2007)
17. Chamoni, P.: BI-Strategie zum Ausgleich von Technologie-Push und Business-Pull, in: Lehner, W.; Piller, G. (eds.): Innovative Unternehmensanwendungen mit In-Memory-Data-Management. Beiträge der Tagung IMDM, pp. 13 – 22 (2011)
18. Gantner, M.: Bereit für Künstliche Intelligenz? is report, Business Intelligence Lösungs-Guide (2018)
19. Hänel, T., Felden, C.: Operational Business Intelligence Meets Manufacturing (2013)
20. Prill, D., Kranzer, S., Merz, R.. Improving Maintenance Processes with Data Science. In: Data Science–Analytics and Applications, pp. 105-107. Springer Vieweg, Wiesbaden (2017)
21. Wiebusch, T., Zimmer, M.: Data Science im Kontext der digitalen Transformation – Nicht nur Data Scientists gesucht, BI-SPEKTRUM 5/2017 (2017)
22. Bundesministerium für Arbeit und Soziales: Grünbuch Arbeiten 4.0 (2015)
23. Heidrich, J.: Data Literacy – Data Skills in der Breite der Hochschulausbildung, in: Gesellschaft für Informatik e.V.: Data Literacy und Data Science Education – Digitale Kompetenzen in der Hochschulausbildung, pp. 6 – 9 (2018)
24. Kirschner, De Bruyckere: The myths of the digital native and the multitasker, Teaching and Teacher Education, Volume 67, pp. 135-142 (2017)
25. Matt, C.; Hess, T.; Benlian, A.: Digital Transformation Strategies, Business and Information Systems Engineering, 57(5), pp. 339–343 (2015)
26. Hoberg, P., Krcmar, H., Welz, B.: Skills for Digital Transformation – Research Report, IDT-Survey (2017)
27. Brenner, W., Broy, M., Leimeister, J.M.: Zur Rolle der Informatik in der Digitalisierung, Diskussionsbeitrag Universität St. Gallen (2017)
28. Demary, V., Klös, H.-P.: Digitalisierung – Kompetenzen für digitale Arbeit, in: Institut der deutschen Wirtschaft Köln (eds.): Perspektive 2035 – Wirtschaftspolitik für Wachstum und Wohlstand in der alternden Gesellschaft, pp. 169 – 181 (2017)
29. Bundesverband Digitaler Wirtschaft: Arbeiten 4.0 – Einfluss der Digitalisierung auf die Arbeitswelt von morgen (2016)
30. Voss, D.: Wie verändert sich der Arbeitsmarkt durch die Digitalisierung?, in: Genderaspekte der Digitalisierung der Arbeitswelt, Arbeitspapier 311 (2018)
31. Zschech, P., Fleißner, V., Baumgärtel, N., Hilbert, A.: Data Science Skills and Enabling Enabling Enterprise Systems – Eine Erhebung von Kompetenzanforderungen und Weiterbildungsangeboten, HMD 55, pp. 163-181 (2018)

Digitaler Konsum: Herausforderungen und Chancen der Verbraucherinformatik

Gunnar Stevens[1], Alexander Boden[2], Lars Winterberg[3],
Jorge Marx Gómez[4], Christian Bala[5]

[1] Universität Siegen, Wirtschaftsinformatik, Siegen, Germany
gunnar.stevens@uni-siegen.de
[2] Fraunhofer-Institut für Angewandte Informationstechnik FIT, Sankt Augustin, Germany
alexander.boden@fit.fraunhofer.de
[3] Universität Regensburg, Vergleichende Kulturwissenschaft, Regensburg, Germany
lars.winterberg@ur.de
[4] Universität Oldenburg, Wirtschaftsinformatik / VLBA, Oldenburg, Germany
jorge.marx.gomez@uni-oldenburg.de
[5] Verbraucherzentrale Nordrhein-Westfalen e. V., Kompetenzzentrum Verbraucherforschung NRW, Düsseldorf, Germany
christian.bala@verbraucherzentrale.nrw

Abstract. Die Durchdringung der Gesellschaft mit IT-Artefakten führt nicht nur zu Veränderungen in der Arbeitswelt („Industrie 4.0“), sondern auch zu einem Wandel in Privathaushalten, etwa im Bereich der Digitalisierung von Verbraucherpraktiken. Dabei werden vor allem die klassischen Konsumfelder Ernährung, Wohnen und Mobilität zunehmend von „smarten“ Geräten und digitalen Diensten durchdrungen. Für VerbraucherInnen eröffnet dies neue Erlebniswelten und vereinfacht den Konsum. Gleichzeitig stellt die neue Qualität der Vernetzung jedoch auch eine Reihe von Fragen, etwa in Bezug auf Datenschutz und die digitale Souveränität der VerbraucherInnen, einen möglichen Digital Divide, sowie die Nachhaltigkeit digital ermöglichter Verbrauchspraktiken auf den Ebenen Gesellschaft, Ökonomie und Ökologie. Die Verbraucherinformatik nimmt dabei ein interdisziplinär verfasstes Forschungsfeld in den Blick, wobei die Bedeutung der Phänomene einer digitalisierten Welt aus Verbrauchersicht untersucht und gestaltet werden sollen.

Keywords: Verbraucherinformatik, Verbraucherforschung, Digitaler Konsum, Nachhaltigkeit, Hauswirtschaft.

1 Einleitung

Die bedarfsgerechte Gestaltung und kompetente Nutzung von Informationssystemen stellt schon lange ein zentrales Forschungsgebiet der Wirtschaftsinformatik dar. Traditionell liegt der Fokus dabei auf *betrieblichen* Informationssystemen. Die fortschreitende Digitalisierung der Gesellschaft reicht mittlerweile jedoch immer stärker auch in den privaten Bereich hinein und beeinflusst damit Verbraucherpraktiken. So werden

14th International Conference on Wirtschaftsinformatik, February 24–27, 2019, Siegen, Germany

z.B. Produktkennzeichnungen nicht mehr allein auf Verpackungen gedruckt, sondern sind vermehrt auch im Internet zugänglich. Bewertungsportale ergänzen traditionelle Ratgeberliteratur, auf denen sich VerbraucherInnen austauschen können. Bargeld wird durch neue Formen bargeldlosen Zahlungsverkehrs ersetzt und das Einkaufen selbst findet im Online-Handel verstärkt digital statt. Neue Haftungsfragen ergeben sich, wenn VerbraucherInnen, statt auf den Rat von BankmitarbeiterInnen zu vertrauen, auch online mit „Robo-Advisorn" interagieren. Entsprechend spielt Digitalisierung bei klassischen Themen der Verbraucherforschung wie Produktsicherheit, Verbraucherrechte, Verbraucherbildung, ethischer und nachhaltiger Konsum etc. eine zunehmend größere Rolle.

Dieser Wandel eröffnet neue Möglichkeiten, die Lebens- und Alltagsökonomien von VerbraucherInnen durch *Smart Services* – verstanden als Kombination von IT- und Dienstleistungs-Design – zu unterstützen. Dies führt jedoch auch dazu, dass VerbraucherInnen im Alltag mit einer zunehmenden Anzahl von Services, Apps, Systemen etc. konfrontiert werden. Durch die konsequente Anwendung von *Usability* und *User-eXperience (UUX)-Methoden* sind einzelne Dienste zwar meist einfach zu nutzen, jedoch entsteht durch die Masse und ihre Wechselwirkungen eine hohe Komplexität, diese zu kombinieren und in den eigenen Alltag zu integrieren. Der Beitrag der Verbraucherinformatik besteht u.a. darin, VerbraucherInnen und ihre Alltagspraktiken als Adressat solcher Systeme ganzheitlich in den Blick zu nehmen.

2 Verbraucherinformatik als interdisziplinäres Forschungsfeld

Der Begriff *Verbraucherinformatik* ist ein Neologismus, der Konnotationen zur Verbraucherforschung wie auch zur Wirtschaftsinformatik aufweist. In Anlehnung an die Definition der Wirtschaftsinformatik kann man die Verbraucherinformatik definieren als *die systematische, methodisch geleitete Untersuchung und Gestaltung von Informations- und Kommunikationstechnologien zur Unterstützung der Haushaltsökonomien und Alltagspraktiken von VerbraucherInnen sowie deren Aneignung und sozialer Einbettung*.

Die Verbraucherinformatik wirbt damit für einen Perspektivwechsel, bei dem nicht nur die Rolle als NutzerInnen eines Systems bzw. KundInnen eines Unternehmens, sondern die individuellen und kollektiven Lebenswelten von VerbraucherInnen möglichst umfassend fokussiert werden. Wie schon in der nicht-digitalen Welt können VerbraucherInnen und ihr Verhalten aus verschiedenen Perspektiven betrachtet werden. Damit ist die Verbraucherinformatik, wie die Verbraucherforschung allgemein [1–3], interdisziplinär konstituiert. Hierbei gehen wir davon aus, dass die digitale Welt nicht ein weiteres Anwendungsfeld der Verbraucherforschung wie Finanzen, Gesundheit, Ernährung, Mobilität etc. darstellt. Vielmehr hat Digitalisierung bereits heute in all diese Felder Einzug gehalten und restrukturiert diese sukzessive. Es stellt also nicht nur einen temporären Trend dar, sondern prägt als ein zentrales Strukturmerkmal des 21. Jahrhunderts weite Teile unseres Lebens – potenziell fortschreitend und zunehmend. Deshalb versteht sich die Verbraucherinformatik als Querschnittsforschung, die sich mit Phänomenen der Digitalisierung in unterschiedlichen Konsumfeldern befasst. Im

Folgenden sollen Anknüpfungspunkte zu bestehenden Forschungsfeldern diskutiert werden, die wichtige Impulse liefern sowie auch Anwendungsbereiche für die Verbraucherinformatik darstellen.

2.1 Die VerbraucherInnen im Blickfeld der angewandten Informatik

In der angewandten Informatik finden sich vielerlei Bezüge. So beschäftigen sich z.B. die Wirtschaftsinformatik, die *Human-Computer-Interaction* (HCI) und die IT-Sicherheitsforschung mit VerbraucherInnen und ihrem Verhalten:

In der **Wirtschaftsinformatik** treten VerbraucherInnen an verschiedenen Stellen in Erscheinung. So untersucht z.B. die wirtschaftsinformatische Nutzerakzeptanzforschung [4], welche Faktoren für die Adaption von IT-Systemen verantwortlich sind. Auf dieser Basis wurden diverse Systeme im Consumermarkt untersucht, wie z.B. Online Shopping [5], Games [6] oder Social Media [7]. In der konstruktiven Wirtschaftsinformatik tauchen VerbraucherInnen z.B. indirekt bei der Gestaltung von Handelsinformationssystemen [8] bzw. direkt bei der Gestaltung von Systemen des *Costumer-Relationship-Managements* (CRM) [9] auf. Beim CRM geht es darum, VerbraucherInnen in ihrer Rolle als KundInnen digital abzubilden sowie die Interkation mit ihnen in die unternehmerischen Prozesse und (IT-)Infrastrukturen einzubetten.

In der **HCI-Forschung** treten VerbraucherInnen primär in der Rolle der NutzerInnen in Erscheinung. Hierbei sind zwei Zweige besonders hervorzuheben: Auf methodischer Ebene ist es das *User Centered Design*, dessen Wurzeln im *Participatory Design* liegen. In den Anfängen ging es diesem primär darum, ArbeiterInnen in die Lage zu versetzen, bei der Gestaltung ihrer Arbeitsplätze mitzuwirken [10]. In der neueren Forschung nehmen demgegenüber die Alltagspraktiken der Menschen außerhalb des Arbeitskontexts eine zunehmend größere Rolle ein. Hierbei werden Ansätze aus der Kulturanthropologie [11], der Kritischen Theorie [12] und den Cultural Studies [13] aufgegriffen, um die Aneignung technischer Artefakte zu verstehen, Konsumpraktiken zu irritieren und vor dem Hintergrund der neuen technischen Möglichkeiten kritisch zu reflektiven [14–16]. Auf theoretischer Ebene sind auch die neueren *User-Experience-Design-Ansätze* hervorzuheben, die in der Bedürfnis- und Konsumpsychologie verankert sind [17, 18], sowie Ansätze des *Privacy By Design* als Reaktion auf die immer tiefergreifende Digitalisierung des Alltags in Kombination mit gängigen Technologien wie dem Erfassen des Nutzerverhaltens zu Werbe- und Optimierungszwecken [19].

In der **Umweltinformatik** [20] wird zunehmend die Bedeutung von VerbraucherInnen für nachhaltiges Wirtschaften (an-)erkannt. Insbesondere wurde die Rückspiegelung des Konsumverhaltens untersucht, um nachhaltiges Verhalten zu befördern [21–23]. Das passende Leitbild stellt dabei das handlungspsychologische Konzept des *Nudging* und des direkten Feedbacks dar, mittels dessen ein umweltbewusstes Handeln motiviert werden soll [24]. Studien zeigen jedoch, dass solche Methoden zwar einen kurzfristigen Effekt haben, jedoch in sich selbst meist nicht nachhaltig sind, weil sie den sozio-materiellen Kontext zu wenig berücksichtigen [25]. Neben Energie werden in der Umweltinformatik auch Konsumfelder wie Ernährung [26] und Mobilität [27] unter einer Nachhaltigkeitsperspektive untersucht. Hier machen z.B. Pakusch et al. [28–30] darauf aufmerksam, dass mit der Einführung autonomen Fahrens nicht nur positive

Nachhaltigkeitseffekte zu erwarten sind – z.B. effiziente Verkehrsflüsse oder der Umstieg vom Auto auf *Shared Autonomous Vehicles* –, sondern auch verschiedene Rebound-Effekte auftreten können. So würden ggf. längere Wegstrecken in Kauf genommen, da sich die Zeit im Auto besser nutzen lässt; ein Umstieg vom ÖPNV auf Shared Autonomous Vehicles ist zu befürchten, da sich Wartezeiten verringern und der Komfort steigt.

In der **IT-Sicherheit** werden VerbraucherInnen als wichtiger Bestandteil sozio-technischer Sicherheitsinfrastruktur betrachtet. Eine maßgebliche Rolle spielt hier auch das sogenannte *Volkszählungsurteil* des Bundesverfassungsgerichts [31], welches die „informationelle Selbstbestimmung" als Teil der Würde des Menschen ausweist. Insbesondere die Forschung zur informierten Einwilligung kann nicht allein technisch gelöst werden, sondern bedarf eines (empirischen) Verständnisses dafür, wie VerbraucherInnen Entscheidungen treffen bzw. inwieweit sie ihre Entscheidungen überblicken können [32–35]. Der Beitrag der IT-Sicherheit für die Verbraucherinformatik besteht darin, ein breites Verständnis zu Sicherheitstechniken und der algorithmischen Analyse großer, heterogener Datensätze (etwa basierend auf dem Tracking von NutzerInnen zu Werbezwecken) zu liefern. Umgekehrt besteht der Beitrag der Verbraucherinformatik für IT-Sicherheit darin, die gesellschaftspolitischen, ökonomischen und sozio-kulturellen Annahmen und Implikationen dieser Mechanismen unter die Lupe zu nehmen und eine evidenzbasierte Grundlage zum Verbraucherhandeln zu liefern (vgl. bspw. Studien zum *Privacy Paradox* [34]). Ferner hilft eine, die politische Ökonomie aufgreifende Verbraucherinformatik dabei, die Ergebnisse gesellschaftspolitisch einzuordnen, um die populäre Fehlannahme zu vermeiden, dass VerbraucherInnen kein Interesse an Datenschutz hätten.

2.2 Die VerbraucherInnen im Blickfeld der Kultur- und Sozialwissenschaften

Neben den dargestellten Kernforschungsfeldern finden sich zusätzliche Anknüpfungspunkte zur existierenden, interdisziplinär verfassten Verbraucherforschung in weiteren Forschungsbereichen, in denen ebenfalls zunehmend die Digitalisierung thematisiert wird:

Die **Rechtswissenschaften** stellen eine wichtige Säule der Verbraucherforschung dar. So wird z.B. VerbraucherIn im Bürgerlichen Gesetzbuch definiert, als „*jede natürliche Person, die ein Rechtsgeschäft zu einem Zwecke abschließt, der weder ihre gewerblichen noch ihrer selbständigen beruflichen Tätigkeit zugerechnet werden kann*" (BGB § 13). Diese Definition bietet analytisch eine gute Grundlage, etwa Phänomene der *Shareconomy* begrifflich zu fassen. Dabei macht es z.B. aus technischer Perspektive keinen Unterschied, ob eine VerbraucherIn ihre Privatwohnung im Urlaub über eine Internetplattform zum Wohnen anbietet oder ob dies durch ein Immobilienunternehmen geschieht. Deshalb wird meist beides vermischt. Aus sozio-kultureller als auch rechtlicher Perspektive handelt es sich jedoch um zwei strikt getrennt zu betrachtende Vorgänge.

Das Kartell- und Wettbewerbsrecht hatte schon immer eine große Bedeutung für den Verbraucherschutz. Insofern ist es nicht verwunderlich, dass Beiträge zur Verbraucherinformatik gerade aus diesem Bereich kommen. Mit der MTS-Kraftstoff-Verordnung

[36] wurden bspw. erstmals digitale Verbraucherinformationsdienste beschrieben und festgelegt: „die Verbraucherinformation, insbesondere die Darstellung, darf nicht irreführend und dadurch geeignet sein, die Entscheidungsfreiheit der Verbraucherinnen und Verbraucher […] zu beeinträchtigen“ [36]. Interessanterweise findet sich kein Pendant hierzu im Bereich der Lebensmittelsicherheit

Die **Psychologie** stellt mit der Werbe- und Wirtschaftspsychologie [37] eine weitere Säule der Verbraucherforschung dar. Deren Erkenntnisse werden insbesondere von der Marketingforschung aufgegriffen, um das Kaufverhalten zu verstehen, zu analysieren und zu gestalten. In neueren Studien lassen sich hierbei zwei sich ergänzende methodologische Strömungen erkennen: Auf der mentalen Ebene versucht z.B. das Neuromarketing die kognitiven Prozesse bei der Bildung von Konsumintentionen präziser zu erfassen [2]. Auf der Verhaltensebene versuchen z.B. neuere *Data Science-Ansätze*, Muster im Verhalten der Masse von VerbraucherInnen zu identifizieren [38]. Daneben postuliert die in der HCI-Forschung aufgegriffene Bedürfnispsychologie [17], dass die Gestaltung der Hauswirtschaft und der Alltagspraktiken nicht allein nach Effizienzkriterien betrachtet werden sollte, sondern dass es den Menschen in der Gesamtheit seiner Bedürfnisse zu betrachten gilt.

Einen weiteren Zweig stellen die **Haushaltsökonomik** und die **Ökotrophologie** dar, die nicht beim einzelnen Subjekt, sondern beim Haushalt bzw. den Alltagspraktiken als primäre Analyseeinheit ansetzen [39, 40]. Der Haushalt wird als Mikrosystem verstanden, das in Makrosysteme wie Gesellschaft und Kultur eingebettet ist, zugleich aber auch ein eigenes intimes Beziehungsgeflecht besitzt [39]. Aus diesem Verständnis analysiert die haushaltswissenschaftliche Verbraucherforschung z.B. Formen der Arbeitsteilungen, etwa bei der Kindererziehung, dem Finanzmanagement, der Verrichtung der Hausarbeit oder der Ernährung [41]. Durch das Smart Home verändert die Digitalisierung zum einen Gegenstände im Haushalt [11, 42, 43], zum anderen verändern sich durch Smart Services die Versorgung des Haushalts mit Produkten und Dienstleistungen [44]. Nicht zuletzt trägt die Digitalisierung mit dazu bei, dass sich Lebensstile und Haushaltsstrukturen selbst wandeln. Hier scheinen sich haushaltsorientierte und verbraucherinformatische Ansätze gegenseitig befruchten zu können.

Die **soziologische, politische** und **kulturanthropologische Verbraucherforschung** macht ferner auf politische, sozio-kulturelle und gesellschaftliche Dimensionen des Konsums aufmerksam [45–48]. Hierzu gehört auch die Frage, wie sich kollektive Interessen von VerbraucherInnen bilden bzw. diese sich (digital) organisieren [49]. Ferner verweisen kulturwissenschaftliche [50, 51] und soziologische Studien [52][53] auf die Historizität und Sozialität von Bedürfnissen. Der jeweilige sozio-ökonomische Kontext ist dabei zentral, so dass Fragen nach der Gestaltung und der Bedürfnisbefriedigung nicht isoliert von den herrschenden Marktmechanismen betrachtet werden sollten [54]. Insbesondere gilt es sich kritisch mit den Wachstumslogiken heutiger Marktwirtschaften auseinanderzusetzen [55], zu deren Prinzipien bspw. das vorzeitige Veralten von Konsumgütern im Sinne zielgerichteter Obsoleszenz [56] genauso gehört wie das Wecken von Bedürfnissen im Allgemeinen [57]: *„Der moderne Verbraucher ist kein Genussmensch, der seinem freien Willen folgt, sondern ein zwanghafter Käufer, zum Konsum getrieben, weil die Zukunft des Kapitalismus davon abhängt“* [57]. Insbesondere die mikroperspektivischen Studien der Kulturanthropologie stehen indes solcherlei deterministischer Zuspitzung mitunter entgegen. So verweisen sie abweichend auch auf die pluralen Logiken alltäglichen Verhaltens, auf individuelle und kollektive Sinnkonstruktionen, die eben nicht zwingend marktförmig sind, sondern sich kapitalistischen Imperativen z.B. auch bewusst oder unbewusst widersetzen, sie symbolisch unterlaufen und bottom-up zu verändern suchen [58–60] .

3 Leitbilder und theoretische Grundlagen

„Das Bild des ‚mündigen Verbrauchers' ist populär in Wissenschaft und Politik. Mündigkeit im Sinne von Kompetenz, Informiertheit und Entscheidungsautonomie ist eine zentrale Annahme herrschender wissenschaftlicher Erklärungsmodelle. Doch diese Modelle werden inzwischen als realitätsfern kritisiert. Von vielen Seiten wird gefordert, realistische Annahmen über das Verbraucherverhalten zu treffen. Denn erst dann könnten Möglichkeiten und Grenzen mündigen Konsums sinnvoll analysiert werden." [61]

Die Anfänge des Konsums als Verbrauch von (Handels-)Gütern findet man spätestens seit der Etablierung des Ackerbaus und der Viehzucht. Treiber dabei waren vor allem die darauf einsetzende Vermögensbildung, die Etablierung des Handels, sowie die zunehmend auf Dauer gestellten Arbeitsteilungen und die Ausbildung von Rechtssystemen [62]. Die modernen VerbraucherInnen heutiger Form existieren aber erst durch die Ausdifferenzierung des modernen Wirtschaftssystems, bei der Produktion, Distribution und Konsumption zunehmend eigenständische Bereiche darstellen. Durch die starke Trennung von Berufs- und Privatleben zeichnete sich das eigene Heim zunehmend auch durch die Funktion aus, die bei der Lohnarbeit verausgabte Arbeitskraft zu regenerieren. Der private Konsum wurde hierbei maßgeblich hinsichtlich der Befriedigung physiologischer Elementarbedürfnisse wie Nahrung, Schlaf, Wärme, etc. verstanden [63]. Mit Einsetzen der Massenkonsum- und Überflussgesellschaft in den westlichen Nationen haben sich vormals elitäre Konsumpraktiken in der Breite etabliert, wobei gleichzeitig Fragen der Freizeitgestaltung, der Mode und des Stils in den Vordergrund rückten [64].

Diese Ausdifferenzierung moderner Konsumgesellschaften bildet damit die Folie einer interdisziplinär verfassten Verbraucherforschung, die sich durch eine hohe Themenvielfalt auszeichnet und zunehmend auch IT-Themen, Daten-Sicherheit und Datenschutz betrachtet [65, 66]. Innerhalb dieser Vielfalt lassen sich aus Sicht der Verbraucherinformatik zunächst drei besonders relevante Strömungen identifizieren: die markttheoretische, konsumtheoretische und praxistheoretische Verbraucherforschung.

3.1 Markttheoretische Verbraucherforschung

Die klassische Verbraucherforschung hat das Markthandeln von VerbraucherInnen zum Gegenstandsbereich. Der Mensch wird dabei als *Homo Oeconomicus* aufgefasst, der jedoch unter Bedingungen mangelnder Markttransparenz und nicht perfekter Märkte agieren muss. Aus den festgestellten Mängeln werden dann meist ordoliberale

Empfehlungen zum Schutz der VerbraucherInnen abgeleitet [67]. Zunehmend werden auch spieltheoretische Modelle zu ungleichen Macht- und Informationsverteilungen zwischen den Marktteilnehmern herangezogen [68] sowie transaktionstheoretische Überlegungen zu den Kosten der Informationsbeschaffung und -verarbeitung [69]. Meist sind VerbraucherInnen aufgrund von Asymmetrien in einer schwächeren Position, woraus sich auch ein besonderer Schutzbedarf ableitet, der seinen Niederschlag z.B. in diversen Verbraucherschutzgesetzen und -verordnungen findet [70]. Es ist deshalb nicht verwunderlich, dass Verbraucherinformationen einen Schwerpunkt der Verbraucherforschung darstellen.

Aus Sicht der VerbraucherInnen lassen sich die Informationen in die drei Kategorien Such-, Erfahrungs- und Vertrauenseigenschaften unterteilen. Diese Informationen können dabei entweder freiwillig bzw. aufgrund rechtlicher Regelungen vom Hersteller stammen oder von Verbraucherinstitutionen und Verbrauchermedien [71]. Die Digitalisierung führt zum einen zu einer Verbreiterung der Quellen und der Kanäle, über die sich VerbraucherInnen über Produkte informieren können. Zum anderen werden die Kosten und der Aufwand zur Informationsbeschaffung gesenkt. So können VerbraucherInnen sich z.B. auf Vergleichsportalen, in Nutzerforen oder über Soziale Medien informieren bzw. sich über ihre Erfahrungen mit Produkten und Dienstleistungen austauschen. Die Kosten der Informationsbeschaffung lassen sich dabei durch Ansätze wie elektronische Kassenzettel weiter senken, da hier VerbraucherInnen nicht jedes einzelne Produkt abscannen müssen, um sich durch Apps wie Barcoo, Codecheck oder fTrace Informationen zum Produkt zu beschaffen [30, 72, 73]. Einen weiteren Trend stellen mobile Verbraucherportale [74] dar. Diese stellen nicht nur Informationen zu Preisen und Eigenschaften von Produkten zur Verfügung, sondern ermöglichen darüber hinaus den Austausch von Erfahrungen und Bewertungen [75].

Offen ist jedoch die Frage, inwieweit und unter welchen Bedingungen es hier zu einer Verschiebung des Kräfteverhältnisses zugunsten der VerbraucherInnen kommt, denn es zeigt sich das zusätzliche Informationen nicht automatisch dazu führen, dass VerbraucherInnen besser informiert sind [76]. So stellt das Positionspapier des *Verbraucherzentrale Bundesverbands* [77] eine Reihe von Forderungen an Vergleichs- und Buchungsportale, um deren Neutralität bewerten zu können. Dazu zählt insbesondere die Gewährung von Transparenz über Geschäftsmodelle, erhaltene Provisionen und den Einfluss und die Gewichtung der Bewertungskriterien bei Produktvergleichen. Hierbei bewegt sich die Wissenschaft stets im Spannungsfeld unterschiedlicher Interessen, wie dies z.B. die Untersuchung zu Vergleichsportalen im Auftrag durch die Verbraucherzentralen [78] als auch im Auftrag von *verivox* [79] exemplarisch zeigt.

In der Verbraucherforschung findet in den letzten Jahren auch eine verstärkte Diskussion zur evidenzbasierten Verbraucherpolitik statt, bei der auch verhaltensökonomische Ansätze aufgegriffen werden [76, 80, 81]. So stehen meist an kognitiv-emotionale Verhaltensweisen ansetzende Maßnahmen im Vordergrund des Interesses, wie z.B. eine verbraucherfreundlichere Regelung von *Opt-in-* und *Opt-out*-Klauseln, welche die Trägheit von VerbraucherInnen berücksichtigen [43]; oder die Gestaltung von Produktinformationsblättern, die die kognitiven Fähigkeiten und emotionalen Verhaltensweisen von VerbraucherInnen beachten [58]. Hierzu gehört explizit auch die Frage nach der visuellen, interaktiven Gestaltung von Verbraucherinformationen und der Aneignung von Informationssystemen durch VerbraucherInnen. Denn die Erkenntnisse der Verhaltensökonomie können sowohl im Sinne der VerbraucherInnen als auch gegen ihr Interesse genutzt werden [82].

3.2 Konsumtheoretische Verbraucherforschung

Konsumtheoretische Ansätze setzen nicht bei der Kaufentscheidung, sondern beim Konsumverhalten an. Belz und Birhatz [70] sprechen hier davon, die klassische Verbraucherforschung um eine zweite Säule zu erweitern. Neben einer verbraucherfreundlichen Marktordnung gilt es auf das Verbraucherverhalten unter dem Leitbild der Nachhaltigkeit einzuwirken. Bei der konsumtheoretischen Forschung wird häufig auf verhaltenspsychologische Handlungstheorien [84] sowie psychologische Bedürfnistheorien [85, 86] Bezug genommen. Der private Konsum wird zum einen in Bezug auf physiologische Elementarbedürfnisse wie Nahrung, Schlaf, Wärme etc. betrachtet [63]. Zum anderen werden die erweiterten, sozialpsychologischen Bedürfnisse und der symbolische Konsum untersucht, der sich in der Massenkonsumgesellschaft in der Breite etabliert hat [64]. Hierbei zeigt sich, dass die Befriedigung der Individualbedürfnisse in der Masse zunehmend in Konflikt mit der ethisch-normativen Forderung einer nachhaltigen Gesellschaft geraten.

Ein Schwerpunkt der neueren Forschung liegt deshalb auch auf der Frage, welche Faktoren für ein nachhaltiges Verhalten verantwortlich sind. Neben utilitaristischen Faktoren, z.B. monetären Anreizen, spielen hier soziale Normen im Sinne der Theorie des sozialen Vergleichens [87] als auch eine persönliche, internalisierte Moral im Sinne der *Norm-Aktivations-Theorie* [88, 89] eine Rolle. Das Ziel verbraucherpolitischer Intervention besteht deshalb nicht nur im Schaffen finanzieller Anreize, sondern auch darin, soziale und persönliche Normen zu aktivieren. Dies geschieht beispielsweise durch Aufzeigen der Folgen (nicht) nachhaltigen, fairen oder solidarischen Handelns, durch Verantwortungszuschreibungen oder Appelle zum Eingreifen. Hierbei ergeben sich enge Bezüge zu den oben genannten *Eco-Feedback*-Systemen und der an VerbraucherInnen orientierten Umweltinformatik.

3.3 Haushalts- bzw. praxistheoretische Verbraucherforschung

Die neueren praxistheoretischen Ansätze [90, 91] setzen an den Haushalts- und Alltagspraktiken von VerbraucherInnen an. Die sozialen Praktiken im eigenen Haushalt sind demnach weniger durch explizite Rollen- und Prozessmodelle geregelt, als durch informelle Absprachen, subtile Aushandlungsprozesse und eingeschliffene Routinen und Gewohnheiten. Dies heißt aber nicht, dass der Haushalt als sozio-materielles Gebilde nicht auch Objekt einer rationalen Gestaltung sein kann, wie dies beispielhaft an der (arbeits-)ökonomisch gestalteten Frankfurter Küche nachvollzogen werden kann [92]. Häufig hat man es aber mit schleichenden, nicht gezielt rational gestalteten Veränderungsprozessen zu tun, wie z.B. der eher organischen Herausbildung der Couchecke [93]. Innerhalb dieses Forschungszweigs wurden unter anderem die Digitalisierung des Haushalts bzw. deren Schwierigkeiten exemplarisch am Beispiel der Gestaltung und Nutzung digitaler Haushaltsbücher untersucht [94]. Die ernüchternden Ergebnisse deuten an, dass verbraucherinformatische Ansätze sich stärker mit der Adaption von IT-Systemen auseinandersetzen müssen, d.h. wie diese angeeignet und in die häuslichen Routinen integriert werden (können) [95]. Einen Schwerpunkt stellen die Ernährungsweisen von VerbraucherInnen dar, die sich auf verschiedenen Ebenen durch die Digitalisierung wandeln, unter anderem durch die Art und Weise wie sich VerbraucherInnen über Lebensmittel und deren Zubereitung mittels digitaler Medien informieren, wie sie Kochen und Essen ästhetisch codieren und digital kommunizieren, wie Produkte über den Online-Handel beschafft werden und wie sich bspw. durch smarte Küchenmaschinen Esskulturen (beschleunigt) wandeln. Denn Haushalt und damit verbundene Lebensweisen werden durch die Digitalisierung selbst transformiert – und damit auch der Gegenstandsbereich der Verbraucherforschung.

In der neueren Verbraucherforschung werden deshalb vermehrt praxistheoretische Ansätze [90] aufgegriffen, um soziale Praktiken des Wohnens, Kochens, Kaufens, Vorsorgens etc. zu analysieren und zu verstehen [91]. Insbesondere werden hier Handlungen nicht als einzelne, diskrete und individuelle Einheiten betrachtet, sondern als sozial eingebettete, situierte, durch implizit-methodisches Wissen geleitete Praktiken verstanden. Sie bilden ein schwach-strukturiertes, dynamisches Geflecht, in dem sich die VerbraucherInnen bzw. der Haushalt als Forschungsgegenstand erst konstituieren [96]. Auf Basis dieses theoretischen Verständnisses haben Ganglbauer et al. [42] gezeigt, dass sich Ernährungsverhalten aus einem Komplex lose gekoppelter Praktiken des Einkaufens, Kochens und Essens zusammensetzt, deren wechselseitige Abhängigkeit es in der Gestaltung von Verbraucherinformationssystemen zu berücksichtigen gilt. Daneben kann man aktuell eine eher schleichende Aneignung intelligenter Einrichtungsgegenstände wie Küchenmaschinen, Staubsaugroboter oder Smart Speakers beobachten. Offen ist, wie diese neuen KI-Materialitäten die häuslichen Routinen und Praktiken verändern werden.

4 Zusammenfassung

„Verbraucherforschung ist Querschnittsforschung und in der Wissenschaft interdisziplinär verortet. Sie findet als nicht eigenständig abgegrenztes Wissensgebiet in den Sozialwissenschaften insbesondere in den Disziplinen der Wirtschaftswissenschaften, der Soziologie, der Politologie und in den Rechtswissenschaften statt wie auch in der psychologischen und medizinischen Forschung.“ [1]

Der kurze Abriss hat gezeigt, dass Verbraucherinformatik als Querschnittsforschung aus Sicht der VerbraucherInnen zu verstehen ist, die sowohl viele Anknüpfungspunkte zu den verschiedenen Gebieten der angewandten Informatik als auch zu den verschiedenen Gebieten einer interdisziplinär verfassten Verbraucherforschung aufweist. Das Ziel des Workshops „Digitaler Konsum – Herausforderungen und Chancen der Verbraucherinformatik“ ist es, verschiedene AkteurInnen und Disziplinen in diesem neuen Feld der Wirtschaftsinformatik zu vernetzen, um ein gemeinsames Verständnis dafür zu entwickeln, was geeignete Themenfelder, Theorien und Methodologien einer pluralen Verbraucherinformatik sein können (vgl. http://www.verbraucherinformatik.de).

Literatur

1. Hagen, K., Oehler, A., Reisch, L.A.: Verbraucherwissenschaft: In welchen Themenbereichen wird geforscht? DIW Wochenbericht. 78, 25–29 (2011).
2. Kenning, P., Oehler, A., Reisch, L.A., Grugel, C.: Verbraucherwissenschaften: Rahmenbedingungen, Forschungsfelder und Institutionen. Springer-Verlag (2017).
3. Nessel, S., Tröger, N., Fridrich, C., Hübner, R.: Multiperspektivische Verbraucherforschung: Ansätze und Perspektiven. Springer-Verlag (2018).
4. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results, (1985).
5. Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: An integrated model. MIS quarterly. 27, 51–90 (2003).
6. Van der Heijden, H.: User acceptance of hedonic information systems. MIS quarterly. 695–704 (2004).
7. Rauniar, R., Rawski, G., Yang, J., Johnson, B.: Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. Journal of Enterprise Information Management. 27, 6–30 (2014).
8. Becker, J., Schütte, R.: Handelsinformationssysteme. MI Wirtschaftsbuch (2004).
9. Ahlert, D., Becker, J., Knackstedt, R., Wunderlich, M.: Customer Relationship Management im Handel: Strategien—Konzepte—Erfahrungen. Springer-Verlag (2012).
10. Ehn, P.: Work-oriented design of computer artifacts, (1988).
11. Crabtree, A., Rodden, T.: Domestic routines and design for the home. Computer Supported Cooperative Work (CSCW). 13, 191–220 (2004).
12. Dunne, A., Raby, F.: Design noir: The secret life of electronic objects. Springer Science & Business Media (2001).
13. Gaver, B., Dunne, T., Pacenti, E.: Design: cultural probes. interactions. 6, 21–29 (1999).
14. Stevens, G., Pipek, V.: Making use: understanding, studying, and supporting appropriation. (2018).
15. Rohde, M., Stevens, G., Brödner, P., Wulf, V.: Towards a paradigmatic shift in IS: designing for social practice. In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. p. 15. ACM (2009).
16. Rohde, M., Brödner, P., Stevens, G., Betz, M., Wulf, V.: Grounded Design–a praxeological IS research perspective. Journal of Information Technology. 32, 163–179 (2017).
17. Diefenbach, S., Hassenzahl, M.: Psychologie in der nutzerzentrieren Produktgestaltung. Springer (2017).
18. Burmester, M., Mast, M., Jäger, K., Homans, H.: Valence method for formative evaluation of user experience. In: Proceedings of the 8th ACM conference on Designing Interactive Systems. pp. 364–367. ACM (2010).
19. Liegl, M., Boden, A., Büscher, M., Oliphant, R., Kerasidou, X.: Designing for ethical innovation: A case study on ELSI co-design in emergency. International Journal of Human-Computer Studies. 95, 80–95 (2016).
20. Teuteberg, F., Gómez, J.M.: Green Computing & Sustainability. HMD Praxis der Wirtschaftsinformatik. 47, 6–17 (2010).

21. Schwartz, T., Denef, S., Stevens, G., Ramirez, L., Wulf, V.: Cultivating energy literacy: results from a longitudinal living lab study of a home energy management system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1193–1202. ACM (2013).
22. Schwartz, T., Stevens, G., Ramirez, L., Wulf, V.: Uncovering practices of making energy consumption accountable: A phenomenological inquiry. ACM Transactions on Computer-Human Interaction (TOCHI). 20, 12 (2013).
23. Schwartz, T., Stevens, G., Jakobi, T., Denef, S., Ramirez, L., Wulf, V., Randall, D.: What people do with consumption feedback: a long-term living lab study of a home energy management system. Interacting with Computers. 27, 551–576 (2015).
24. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1999–2008. ACM (2010).
25. Gram-Hanssen, K.: Understanding change and continuity in residential energy consumption. Journal of Consumer Culture. 11, 61–78 (2011).
26. Clear, A.K., Comber, R., Friday, A., Ganglbauer, E., Hazas, M., Rogers, Y.: Green food technology: UbiComp opportunities for reducing the environmental impacts of food. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. pp. 553–558. ACM (2013).
27. vom Berg, B.W., Köster, F., Gómez, J.M.: Elektromobilität: Gegenwart oder Zukunft? Multikonferenz Wirtschaftsinformatik 2010. 195 (2010).
28. Pakusch, C., Bossauer, P., Shakoor, M., Stevens, G.: Using, Sharing, and Owning Smart Cars. In: Proceedings of the 13th International Joint Conference on e-Business and Telecommunications. pp. 19–30. SCITEPRESS-Science and Technology Publications, Lda (2016).
29. Pakusch, C., Stevens, G., Boden, A., Bossauer, P.: Unintended effects of autonomous driving: A study on mobility preferences in the future. Sustainability. 10, 2404 (2018).
30. Vonholdt, S., Stevens, G., Kleith, K., Boden, A.: Digitale Rückverfolgbarkeit von Lebensmitteln: Eine verbraucherinformatische Studie. Proceedings of the Tagungsband Wirtschaftsinformatik. (2019).
31. Bundesverfassungsgericht: BVerfGE 39, 1. BVerfGE. 88, 203–218 (1983).
32. Jakobi, T., Stevens, G., Castelli, N., Ogonowski, C., Schaub, F., Vindice, N., Randall, D., Tolmie, P., Wulf, V.: Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2, 171 (2018).
33. Krasnova, H., Spiekermann, S., Koroleva, K., Hildebrand, T.: Online social networks: Why we disclose. Journal of Information Technology. 25, 109–125 (2010).
34. Palen, L., Dourish, P.: Unpacking privacy for a networked world. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 129–136. ACM (2003).
35. Stevens, G., Wulf, V.: Computer-supported access control. ACM Transactions on Computer-Human Interaction (TOCHI). 16, 12 (2009).
36. BMWi: MTS-Kraftstoff-Verordnung vom 22. März 2013 ärz 2013 (BGBl. I S. 595, 3245; 2013 I S. 3304). (2013).
37. Moser, K.: Wirtschaftspsychologie. Springer (2007).
38. Erevelles, S., Fukawa, N., Swayne, L.: Big Data consumer analytics and the transformation of marketing. Journal of Business Research. 69, 897–904 (2016).
39. Gräbe, S.: Der private Haushalt im wissenschaftlichen Diskurs. Campus-Verlag (1993).
40. Piorkowsky, M.-B.: Konsum aus Sicht der Haushaltsökonomik. In: Konsum. pp. 187–206. Springer (2000).

41. Seel, B.: Haushaltsökonomik - Grenzgänge zwischen Wissenschaftsparadigmen: Forschungszentrum Innovation und Dienstleistung. 54, 112–119 (2006).
42. Ganglbauer, E., Fitzpatrick, G., Comber, R.: Negotiating food waste: Using a practice lens to inform design. ACM Transactions on Computer-Human Interaction (TOCHI). 20, 11 (2013).
43. Jakobi, T., Ogonowski, C., Castelli, N., Stevens, G., Wulf, V.: The catch (es) with smart home: Experiences of a living lab field study. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 1620–1633. ACM (2017).
44. Heinemann, G.: Der neue Online-Handel. Springer (2014).
45. Baringhorst, S., Kneip, V., März, A., Niesyto, J.: Politik mit dem Einkaufswagen. Unternehmen und Konsumenten als Bürger in der globalen Mediengesellschaft. Bielefeld: transcript. (2007).
46. Baringhorst, S., Witterhold, K.: Zwischen Markt und Zivilgesellschaft–Organisation von Verbraucherinteressen Online/Offline. In: Verbraucherwissenschaften. pp. 557–572. Springer (2017).
47. Hirschfelder, G., Ploeger, A., Rückert-John, J., Schönberger, G.: Was der Mensch essen darf: Ökonomischer Zwang, ökologisches Gewissen und globale Konflikte. Springer-Verlag (2014).
48. Winterberg, L.: Die Not der Anderen: Kulturwissenschaftliche Perspektiven auf Aushandlungen globaler Armut am Beispiel des Fairen Handels. Bausteine einer Ethnografie. Waxmann Verlag (2017).
49. Strünck, C.: Politikwissenschaftliche Perspektive. In: Verbraucherwissenschaften. pp. 123–140. Springer (2017).
50. Hirschfelder, G.: Europäische Esskultur: Eine Geschichte der Ernährung von der Steinzeit bis heute. Campus Verlag, Frankfurt/Main (2005).
51. Hirschfelder, G.: Facetten einer Ernährungs-Globalgeschichte. Esskultur als Resultat historischer Prozesse, http://www.bpb.de/apuz/262254/ernaehrungs-globalgeschichte.
52. Peters, C.H., Schulz, P.: Resonanzen und Dissonanzen: Hartmut Rosas kritische Theorie in der Diskussion. transcript Verlag (2017).
53. Shove, E.: Comfort, cleanliness and convenience: The social organization of normality (New Technologies/New Cultures). (2004).
54. König, G.M.: Konsumkultur: Inszenierte Warenwelt um 1900. Böhlau, Wien (2007).
55. Paech, N.: Befreiung vom Überfluss. oekom-Verlag (2012).
56. Hillmann, K.-H.: Geplante Obsoleszenz. Bemerkungen zu Burkhardt Röper:» Gibt es geplanten Verschleiß?«. Zeitschrift für Verbraucherpolitik. 1, 48–61 (1977).
57. Barber, B.R.: Consumed!: wie der Markt Kinder verführt, Erwachsene infantilisiert und die Bürger verschlingt. CH Beck (2007).
58. Klein, I., Windmüller, S.: Kultur der Ökonomie: Zur Materialität und Performanz des Wirtschaftlichen. transcript, Bielefeld (2014).
59. Langreiter, N., Löffler, K.: Selber machen: Diskurse und Praktiken des »Do it yourself«. transcript, Bielefeld (2017).
60. Tauschek, M., Grewe, M.: Knappheit, Mangel, Überfluss: Kulturwissenschaftliche Positionen zum Umgang mit begrenzten Ressourcen. Campus Verlag, Frankfurt (2015).
61. Strünck, C., Arens-Azevêdo, U., Brönneke, T., Hagen, K., Jaquemoth, M., Kenning, P., Liedtke, C., Oehler, A., Tamm, M.: Ist der „mündige Verbraucher "ein Mythos? Auf dem Weg zu einer realistischen Verbraucherpolitik. Stellungnahme des Wissenschaftlichen Beirats Verbraucher- und Ernährungspolitik beim BMELV (2012).
62. Reichholf, J.H.: Warum die Menschen sesshaft wurden: Das größte Rätsel unserer Geschichte. S. Fischer Verlag (2010).

63. Eder, F.X.: Geschichte des Konsumierens–Ansätze und Perspektiven der (historischen) Konsumforschung. Konsumieren in Österreich–19. und. 20, 9–41 (2006).
64. Bocock, R.: Consumption. Routledge (2008).
65. Bala, C., Müller, K.: Abschied vom Otto Normalverbraucher: Moderne Verbraucherforschung: Leitbilder, Information, Demokratie. Klartext, Essen (2015).
66. Bala, C., Müller, K. eds: Der gläserne Verbraucher: Wird Datenschutz zum Verbraucherschutz? Verbraucherzentrale NRW, Düsseldorf (2014).
67. Kollmann, K.: Eingeschränkt–statt ermächtigt. Journal für Verbraucherschutz und Lebensmittelsicherheit. 7, 393–395 (2012).
68. Akerlof, G.A.: The market for" lemons": Quality uncertainty and the market mechanism. The quarterly journal of economics. 488–500 (1970).
69. Williamson, O.E.: Transaction Cost Economics. In: Menard, C. and Shirley, M.M. (eds.) Handbook of New Institutional Economics. pp. 41–65. Springer US (2005).
70. Belz, F.-M., Bilharz, M.: Nachhaltiger Konsum, geteilte Verantwortung und Verbraucherpolitik: Grundlagen. Nachhaltiger Konsum und Verbraucherpolitik im. 21, 21–52 (2007).
71. Schoenheit, I.: Was Verbraucher wissen wollen. Studienreport (2004).
72. Bossauer, P., Hanschke, S., Stevens, G.: Mehrwerte auf Basis digitaler Kassenzettel: eine verbraucherinformatische Studie. In: Nachhaltiges Wirtschaften im digitalen Zeitalter. pp. 135–150. Springer (2018).
73. Stevens, G., Bossauer, P., Neifer, T., Hanschke, S.: Using shopping data to design sustainable consumer apps. In: 2017 Sustainable Internet and ICT for Sustainability (SustainIT). pp. 1–3. IEEE (2017).
74. Loos, D.J., Braun, D.: Die neue Macht der Konsumenten? Einfluss von Verbraucherportalen auf die Verpackungsgestaltung im Lebensmittelbereich. In: Hoffmann, S., Schwarz, U., and Mai, R. (eds.) Angewandtes Gesundheitsmarketing. pp. 177–190. Springer Fachmedien Wiesbaden (2012).
75. Thiem, H.-M.: Externe Kommunikation im Unternehmen durch Social Media: Marketing, Networking, Recruiting und Marktforschung. Diplomarbeiten Agentur (2012).
76. Strünck, C.: Die Verbraucherpolitik braucht Pragmatismus statt wirklichkeitsferner Leitbilder. Wirtschaftsdienst. 91, 165–168 (2011).
77. VZBV: Vergleichs- und Buchungsportale. Positionspapier des Verbraucherzentrale Bundesverband (2017).
78. Dautzenberg, K., Grasl, K., Halm, T., Schulze, A.: Buchungs- und Vergleichs-Portale – Die schwierige Suche nach dem günstigsten Preis. Untersuchung der Verbraucherzentralen im Rahmen des Projekts Marktwächter Digitale Welt (2016).
79. Knieper, T.: Gutachten zur Untersuchung der Verbraucherzentralen „Buchungs- und Vergleichsportale – Die schwierige Suche nach dem günstigsten Preis“. Gutachten im Auftrag von verivox (2016).
80. Hagen, K., Micklitz, H.-W., Oehler, A., Reisch, L.A., Strünck, C.: „Check Verbraucherpolitik und Verbraucherbeteiligung “–Empfehlungen für eine evidenzbasierte Verbraucherpolitik. Journal für Verbraucherschutz und Lebensmittelsicherheit. 1–6 (2013).
81. Kenning, P., Wobker, I.: Ist der „mündige Verbraucher “eine Fiktion? Ein kritischer Beitrag zum aktuellen Stand der Diskussion um das Verbraucherleitbild in den Wirtschaftswissenschaften und der Wirtschaftspolitik. Zeitschrift fuer Wirtschafts-und Unternehmensethik. 14, (2013).
82. Oehler, A., Reisch, L.A.: Behavioral Economics–eine neue Grundlage für die Verbraucherpolitik. Studie im Auftrag des vzbv eV, Berlin. (2008).

83. Tiffe, A., Feigl, M., Fritze, J., Götz, V., Grunert, C., Jaroszek, L., Rohn, I.: Ausgestaltung eines Produktinformationsblatts für zertifizierte Altersvorsorge-und Basisrentenverträge. (2012).
84. Ajzen, I., Fishbein, M.: Understanding attitudes and predicting social behaviour. (1980).
85. Maslow, A.H.: A theory of human motivation. Psychological review. 50, 370 (1943).
86. Sheldon, K.M., Elliot, A.J., Kim, Y., Kasser, T.: What is satisfying about satisfying events? Testing 10 candidate psychological needs. Journal of personality and social psychology. 80, 325 (2001).
87. Festinger, L.: A theory of social comparison processes. Human relations. 7, 117–140 (1954).
88. Schwartz, S.H.: Normative influences on altruism1. In: Advances in experimental social psychology. pp. 221–279. Elsevier (1977).
89. Stern, P.C., Kalof, L., Dietz, T., Guagnano, G.A.: Values, beliefs, and proenvironmental action: Attitude formation toward emergent attitude objects 1. Journal of applied social psychology. 25, 1611–1636 (1995).
90. Reckwitz, A.: Toward a theory of social practices: A development in culturalist theorizing. European journal of social theory. 5, 243–263 (2002).
91. Warde, A.: Consumption and theories of practice. Journal of consumer culture. 5, 131–153 (2005).
92. Kramer, L.: Rationalisierung des Haushaltes und Frauenfrage - Die Frankfurter Küche und Zeitgenössische Kritik. Ernst May und das neue Frankfurt, 1925-1930 (Berlin, Verlag für Architektur und Technische Wissenschaften, 1986), pp. 77Á84. (1925).
93. Warnke, M.: Zur Situation der Couchecke. Stichworte zur ‚Geistigen Situation der Zeit', Frankfurt a. M. 673–687 (1979).
94. Piorkowsky, M.-B.: Household accounting in Germany–Some statistical evidence and the development of new systems. Accounting, Auditing & Accountability Journal. 13, 518–534 (2000).
95. Silverstone, R., Haddon, L.: Design and the domestication of information and communication technologies: Technical change and everyday life. (1996).
96. Halkier, B., Katz-Gerro, T., Martens, L.: Applying practice theory to the study of consumption. Journal of Consumer Culture. 11, 3–13 (2011).

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.dnb.de** abrufbar.

Impressum

Herausgeber: Thomas Ludwig, Volkmar Pipek

Siegen 2019: univer*si* – Universitätsverlag Siegen
www.uni-siegen.de/universi

ISBN 978-3-96182-063-4